AD NUMBER

AD828731

LIMITATION CHANGES

TO:

Approved for public release; distribution is unlimited.

FROM:

Distribution authorized to U.S. Gov't. agencies and their contractors; Critical Technology; JAN 1968. Other requests shall be referred to Air Force Technical Application Center, Washington, DC 20333. This document contains export-controlled technical data.

AUTHORITY

usaf ltr, 28 feb 1972

DESIGN AND EVALUATION OF CERTAIN MULTICHANNEL FILTERS

26 January 1968

Prepared for

AIR FORCE TECHNICAL APPLICATIONS CENTER
Washington, D. C.

By

D. W. McCowan

TELEDYNE, INC.

Under

Project VELA UNIFORM

Sponsored By

ADVANCED RESEARCH PROJECTS AGENCY
Nuclear Test Detection Office
ARPA ORDER NO. 624

DESIGN AND EVALUATION OF CERTAIN MULTICHANNEL FILTERS

SEISMIC DATA LABORATORY REPORT NO. 209

AFTAC Project No.:	VELA T/6702
Project Title:	Seismic Data Laboratory
ARPA Order No.:	624
ARPA Program Code No.:	5810
Name of Contractor:	TELEDYNE, INC.
Contract No.:	F 33657-67-C-1313
Date of Contract:	2 March 1967
Amount of Contract:	$ 1,736,617
Contract Expiration Date:	1 March 1968
Project Manager:	William C. Dean (703) 836-7644

P.O. Box 334, Alexandria, Virginia

AVAILABILITY

This document is subject to special export controls and each transmittal to foreign governments or foreign nationals may be made only with prior approval of Chief, AFTAC.

Washington

This research was supported by the Advanced Research Projects Agency, Nuclear Test Detection Office, under Project VELA-UNIFORM and accomplished under the technical direction of the Air Force Technical Applications Center under Contract F 33657-67-C-1313.

Neither the Advanced Research Projects Agency nor the Air Force Technical Applications Center will be responsibile for information contained herein which may have been supplied by other organizations or contractors, and this document is subject to later revision as may be necessary.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

The theory of optimum least-squares multichannel filters is developed from the definitions of correlation functions and power spectra. Four types of these optimum processors were applied to an Aleutian earthquake recorded by part of the Montana LASA and the results evaluated. It appears that certain isotropic processors can give a modest signal-to-noise ratio improvement in excess of 1 db over the prefiltered straight sum. Evaluation of the outputs and transfer functions of these filters gives some insight into methods for improving their gain and stability. Several other topics are discussed from a theoretical point of view; these include estimation of power spectral density functions, frequency-wavenumber spectra, and certain coherence functions.

INTRODUCTION

This report concludes the first phase of the SDL multichannel filter project; namely, the design and evaluation of certain multichannel filters. It has two parts: (1) the theoretical specification and formal solution of the digital filtering problem; and (2) the experimental results obtained by applying the various multichannel filters to actual seismic array data. On this basis certain conclusions have been drawn about the performance of the various multichannel filters.

The practical computation of many of these filter operators relies heavily on the fast Fourier transform algorithm of Cooley and Tukey, without which many of the calculations reported here would be too expensive to be practical and some would not even be possible. An earlier report (McCowan, 1966) dealt with the transform method and its associated techniques and the present report can be regarded as a logical development of the ideas presented there.

Throughout this report the standard of comparison used to grade multichannel filter performance is the unbeamsteered sum of the data channels. This was chosen instead of the beamsteered or phased sum simply as a matter of convenience. Straight summation is itself an optimum processor when the signals are aligned and the noise is uncorrelated. Any departures from this model either in signal misalignment or noise correlation, will, of course, detract from the theoretically best performance, which is the square root of the number of sensors increase in signal to noise (S/N) ratio improvement measured in decibels (db). This is where the multichannel filter is applicable; the design equations for which are a theoretical statement of these departures in signal and noise models. Filters satisfying these equations should provide an increase in S/N ratio improvement over the otherwise optimum processor, the straight sum.

The data used in the experimental part of this report were from an Aleutian earthquake recorded on ten sensors in the D1 subarray of the Montana LASA. The pertinent facts about the event are listed in Table 1. Ten sensors were used instead of the original twenty-five available in the D1 subarray to reduce the computation time and to simplify the procedure of tabulating results. Calibrations were computed in the usual manner and all traces were corrected to true ground motion. Table 2 lists the coordinates and calibrations (demagnifications) for the set of ten sensors. This particular set was selected to provide a reasonably geometric array with good element separation. Figure 1 is a map of the reduced array.

For some parts of the study, the traces were prefiltered with the "SDL" bandpass filter. This filter is a phaseless four-pole Butterworth recursive filter. It was chosen to provide a balance between excessive signal distortion at one end and good noise reduction at the other. The frequency response of this filter is shown in Figure 2.

As in previous SDL data-processing reports, the definition of signal amplitude will be one-half the maximum peak-to-trough amplitude occurring in the first three cycles of the P phase. Noise is defined as the root-mean-square amplitude measured over some specified interval. The S/N ratio is the quotient of these two quantities. Frequently results will be expressed in db for which the usual expression will be used, $20 \log_{10}(\text{RATIO})$. When the quantities to be expressed in terms of db have the dimensions of power, this definition has been changed to $10 \log_{10}(\text{RATIO})$ in order to preserve the proper amplitude ratio.

Finally, much information in this report is presented in contour plots. These plots are drawn by the computer on the printer. Accuracy and resolution in these plots are balanced against the time necessary to compute and print

them. As a result, sometimes the fine detail predicted by the theory is lost. Nevertheless, printer contour plots have been used unretouched because they are an efficient representation of large amounts of information. For each plot we computed the function to be plotted only at a grid of 21 x 21 evenly spaced points in the horizontal and vertical directions, and fitted in a grid of 60 x 100 points by bilinear interpolation.

A. MEASUREMENT OF SEISMIC NOISE

The design of digital filters and the analysis of their inputs and outputs require the estimation of such quantities as correlation functions, power spectral density functions, frequency-wavenumber spectra and various coherence functions. These measurements are the connections between theoretical derivations and actual practice. In this section we will describe the methods we have selected for computing these functions as well as any consequences these methods may imply on the accuracy or form of the results.

1. Correlation Functions. In the continuous case a general correlation function is defined by an integral similar to a convolution:

$$R_{ij}(t) = \frac{1}{T} \int_0^T X_i(\tau)\, X_j(t+\tau)\, d\tau \qquad (1)$$

When i=j, R is an autocorrelation and when i≠j, it is a cross-correlation. In the discrete case the integral becomes a sum:

$$R_{ij}(t) = \sum_{\tau} X_i(\tau)\, X_j(t+\tau) \qquad (2)$$

where t and τ are now indices. In the work that follows they will still be written as arguments to avoid confusion.

It is usually assumed that the time series under consideration are zero outside the region of interest. This has the

result of bounding the correlation function and ensuring that it tends towards zero as 100% lags, the number of lags equal to the number of points in the time series, are taken. This condition can be written explicitly by limiting the sum in equation 2:

$$R_{ij}(t) = \sum_{\tau=0}^{T-t-1} X_i(\tau)\, X_j(t+\tau) \qquad (3)$$

Negative lags are taken care of by use of the identity:

$$R_{ij}(t) = R_{ji}(-t) \qquad (4)$$

One further generalization is necessary to take into account stationarity. If the time series under consideration are stationary in time then it does not matter what sample is used (provided it is long enough to ensure stability) to estimate the correlation functions. That is:

$$R_{ij}(t) = \sum_{\tau=s}^{T-t+1+s} X_i(\tau)\, X_j(t+\tau) \qquad (5)$$

for all s. This will be needed later in the least-squares derivation.

2. Power Spectral Density Functions. Prior to the development of the fast Fourier transform method, power spectra were invaribly computed by transforming the respective correlation function. This method is usually referred to as the Blackman-Tukey (B-T) method of spectral estimation (Blackman and Tukey, 1959). However, since computing Fourier transforms of long pieces of data directly is now economical in terms of machine time and storage space, a new method has come into vogue. We have chosen to compute the spectra used in this project by direct multiplication of Fourier transforms. This procedure is now referred to as the Cooley-Tukey (C-T) method. (Cooley and Tukey, 1965; Cooley et. al., 1967; McCowan, 1966).

The starting point in the B-T method is equation (2). However, because the computation of correlation functions is costly in terms of machine time, those generally used are not 100% lag correlation functions. The effect of their truncation is to introduce the spectral window. Multiplication by a "boxcar" function in time is equivalent to convolution by its transform, the $\frac{\sin x}{x}$ function, in frequency. Thus the result of the B-T method is the 100% lag power spectra smoothed by this spectral window. The more lags that are used, the narrower is the spectral window. Therefore the correlation functions are computed out only as far as required to give a balance between stability and accuracy on the one hand and economy in computing time on the other.

The C-T method and the necessary smoothing that is chosen produce a different set of constraints and criteria to be optimized. The method is superior only when the length of the time series is a highly composite number. In particular our program works only when the number of points in the time series is a power of two. When this is so the C-T method produces a N+1 point raw power spectral density function, where $N = 2^L$ is the length of the data. Since N is of the order of 4096 or 2048, this represents a large amount of information. In fact these N+1 points are the actual transform of the 100% lag correlation function. Because a set of auto and cross spectra for 13 or 21 channels of data each with 4097 frequencies between dc and the folding frequency greatly exceeds the capacity of our machine, it is reasonable to sacrifice some of these degrees of freedom to smooth and to reduce the spectra to a more manageable size. Therefore we have chosen to apply a smoothing and decimation procedure several times to reduce the raw spectra to something like 65 or 129 points each.

Our procedure consists of convolving the power spectra with the three point operator, $(\frac{1}{4},\frac{1}{2},\frac{1}{4})$, and then taking every

other point of the result. To see what this does to the 100% lag correlation function we again use the convolution theorem. Convolution by (¼,½,¼) in frequency is equivalent to multiplication by its Fourier transform in time which is:

$$F(k) = \tfrac{1}{2} + \tfrac{1}{4}e^{\frac{2\pi i k}{2N}} + \tfrac{1}{4}e^{\frac{-2\pi i k}{2N}} = \tfrac{1}{2}\left(1+\cos\frac{\pi k}{N}\right) = \cos^2\left(\frac{\pi k}{2N}\right) \tag{6}$$

The 100% lag correlation function, which includes N-1 points on either side of the zero lag value, is therefore multiplied by this function or window. Taking every other point of the result cuts in half each side of the windowed correlation function. A good measure of howmuch error is introduced by L of these procedures is the relative energy folded back into the lower lags of the correlation function.

$$E_L = \left[\frac{\int_{T/2}^{T} \cos^2 \frac{\pi t}{2T}\, dt}{\int_{0}^{T} \cos^2 \frac{\pi t}{2T}\, dt}\right]^L = \left(\tfrac{1}{2} - \frac{1}{\pi}\right)^L \tag{7}$$

The advantages of the C-T method are numerous; but, the two major features are its inherent speed when used in conjunction with the fast Fourier transform and the unrestricted choice of smoothing it permits. For reasonable choices of smoothing it cannot fail to be more accurate because weighted information from all lags of the correlation function appears in the spectra. We have noticed that an effect of this is to give more detail to the spectra, for equivalent degrees of freedom, while still following closely the overall behavior of B-T spectra.

3. Frequency-Wavenumber Spectra. The spatial information of an array of seismometers can be used in conjunction with their frequency spectra to yield velocity distribution

information about the data. This technique consists of computing frequency-wavenumber spectra and then viewing the resulting plots in polar coordinates. The distance from the origin is related to the velocity by:

$$|\bar{v}| = \frac{f}{|\bar{k}|} \tag{8}$$

Our convention on direction is that the arrow pointing from the origin to the noise peak on the diagram is opposite in direction to the propagation vector of the data. In other words, the f-k plots in this report represent the direction from which the data is coming rather than the direction toward which it is propagating.

Just as regular power spectra are the squares of a time transform, f-k spectra are the squares of a space transform. In the continuous case this is:

$$P(f,\bar{k}) = \left| \int dt \int \overline{dx}\, f(t,\bar{x})\, \exp[-2\pi i(ft-\bar{k}\cdot\bar{x})] \right|^2 \tag{9}$$

where for convenience we have written the set of data as $f(t,\bar{x})$.

Expanding the square gives:

$$P(f,\bar{k}) = \int dt_1 \int d\bar{x}_1\, f(t_1,\bar{x}_1)\, \exp[2\pi i(ft_1-\bar{k}\cdot\bar{x}_1)]$$

$$\int dt_2 \int d\bar{x}_2\, f(t_2,\bar{x}_2)\, \exp[-2\pi i(ft_2-\bar{k}\cdot\bar{x}_2)] \tag{10}$$

Now changing to new variables $q = t_2-t_1$ and $\bar{s} = \bar{x}_2-\bar{x}_1$

$$P(f,\bar{k}) = \int dq \int ds \left[\int dt_1 \int d\bar{x}_1\, f(t_1,\bar{x}_1)\, f(q+t_1,\bar{s}+\bar{x}_1) \right]$$

$$\exp[-2\pi i(fq-\bar{k}\cdot\bar{s})] \tag{11}$$

The bracketed quantity is called the time-space correlation function. Equation (11) states that an f-k spectrum is the

time-space Fourier transform of a time-space correlation function.

However, this expression simplifies for the case when the data $f(t,\bar{x})$ are sampled at the set of coordinates $\{\bar{x}_i\}$. In this case $f(t,\bar{x})$ can be written:

$$f(t,\bar{x}) = \sum_i f_i(t)\ \delta(\bar{x}-\bar{x}_i) \tag{12}$$

Substituting this and integrating over $\bar{x}$, gives:

$$P(f,\bar{k}) = \int dq \int d\bar{s} \sum_i \sum_j \left[\int dt, f_i(t_1)\ f_j(q+t_1)\delta(\bar{x}_i-\bar{x}_j+\bar{s})\right] \exp[-2\pi i(fq-\bar{k}\cdot\bar{s})] \tag{13}$$

Inserting the definition of the correlation functions, equation (1), and integrating over s gives:

$$P(f,\bar{k}) = \sum_i \sum_j \left[\int dq\ R_{ij}(q)\ \exp[-2\pi i f q]\right] \exp[+2\pi i\bar{k}\cdot(\bar{x}_i-\bar{x}_j)] \tag{14}$$

The bracketed quantities here are the spectral matrix elements, in terms of which the final result can be written:

$$P(f,\bar{k}) = \sum_i \sum_j S_{ij}(f)\ \exp[+2\pi i\bar{k}\cdot(\bar{x}_i-\bar{x}_j)] \tag{15}$$

This formula is also applicable to the calculation of filter f-k responses where again the time-space function is defined only at a given set of coordinates.

Inherent in this formalism is a window. Just as transforming a regular correlation function gave a frequency window, so transforming a spatial correlation function will give a k-space window. A representation of this window is the array response function which is analogous to the transfer function

of a filter. To obtain this we put a time impulse into each channel at the same instant of time. Since the spectral matrix of a set of impulses is constant:

$$S_{ij}(f) = 1 \tag{16}$$

the array response is given by:

$$R(\bar{k}) = \sum_i \sum_j \exp[+2\pi i \bar{k} \cdot (\bar{x}_i - \bar{x}_j)] \tag{17}$$

Any data displayed in an f-k spectrum plot computed by equation (15) will be seen as the actual f-k spectrum multiplied by the array response function. A further use of the array response is as a criterion in optimizing the array geometry; however, in this report, we will be concerned only with given array geometries.

4. Coherence Functions. A useful measure of the correlation in frequency or coherence of two data channels is the normalized magnitude of the cross spectrum. This is called an ordinary coherence function:

$$\gamma^2_{ij}(\omega) = \frac{|S_{ij}(\omega)|^2}{S_{ii}(\omega)\, S_{jj}(\omega)} \tag{18}$$

An ordinary coherence function is a result of the smoothing applied to the spectra. This is because, for no smoothing:

$$\begin{aligned} S_{ii}(\omega) &= |F_i(\omega)|^2 \\ S_{ij}(\omega) &= F_i^*(\omega)\, F_j(\omega) \end{aligned} \tag{19}$$

and therefore in this case:

$$\gamma^2_{ij}(\omega) = 1 \tag{20}$$

However when expected values are taken the ordinary coherence takes on its customary meaning. The matrix of ordinary coherence functions follows from the definition but is usually referred to as the normalized spectral matrix.

Two other coherence functions are the multiple and partial coherences. It is only possible to define them within the framework of least-squares filter theory. This is done in the appendix. At this point it is sufficient to say that the multiple coherence between a given data channel and a set of other channels is a measure of the amount of power in that channel that can be represented by a linear combination of the other channels, and that the partial coherence between two given data channels with the effects of a set of other channels removed is the ordinary coherence of those two channels with the others prediction-filtered out.

B. Multichannel Least-Squares Filters

With the exception of the maximum-likelihood filter, all of the filters described in this report are solutions of the Wiener-Hopf equation. The maximum-likelihood filter itself is a solution of this equation in a different form. However, all of these filters are multichannel filters in that they satisfy the multichannel convolution algorithm:

$$y(t) = \sum_i \sum_\tau f_i(\tau)\, x_i(t-\tau)$$

$$y(\omega) = \sum_i f_i(\omega)\, x_i(\omega) \qquad (21)$$

in the time and frequency domains respectively. Here y is the output, f_i is the ith filter, and x_i is the ith input data channel. In this section we will derive the Wiener-Hopf equation in both the time and frequency domains and show its application to the multichannel filters under consideration. We will also only be concerned with the discrete case

primarily because of the simplification inherent in it.

1. The Wiener-Hopf Equation. The derivation we will give of the Wiener-Hopf equation is actually a derivation of the general wave-shaping filter. This we will do first for the time-domain case by making use of equation (21) to write an expression for the error in energy between the output of a multichannel filter and a desired output d(t). This is:

$$E = \sum_t [y(t)-d(t)]^2 = \sum_{ti\tau} f_i(\tau)\, X_i(t-\tau) \sum_{js} f_j(s)\, X_j(t-s) \qquad (22)$$

$$-2\sum_t d(t) \sum_i f_i(\tau)\, X_i(t-\tau) + \sum_t d(t)^2$$

With the substitution $q = t-\tau$ and using the definition of correlation, equation (3), the first terms can be re-written as:

$$E = \sum_{ijts} R_{ij}(\tau-s)\, f_i(\tau)\, f_j(s) - 2\sum_{i\tau} R_{id}(\tau)\, f_i(\tau) + \sum_t d(t)^2 \qquad (23)$$

This step also implies stationarity, equation (5). E is now considered to be a function of the filter coefficients $f_i(t)$ which are to be varied in order to produce a minimum value of E. This is accomplished by setting the derivatives of E with respect to the coefficient equal to zero:

$$\frac{\partial E}{\partial f_k(p)} = 0 = 2 \sum_{i\tau} R_{ki}(p-\tau)\, f_i(\tau) - 2\, R_{kd}(p) \qquad (24)$$

The result is the normal equations:

$$\sum_{\tau i} R_{ki}(p-\tau)\, f_i(\tau) = R_{kd} \qquad (25)$$

These, written in matrix notation, become:

$$R_{xx}F = R_{xd} \qquad (26)$$

Where R_{xx} is a very large matrix composed of smaller matrices in the form:

$$R_{xx} = \begin{pmatrix} r_{xx}(0) & r_{xx}(-1) & r_{xx}(-2) & \cdots & r_{xx}(-L+1) \\ r_{xx}(1) & r_{xx}(0) & r_{xx}(-1) & & \vdots \\ r_{xx}(2) & r_{xx}(1) & r_{xx}(0) & & \vdots \\ \cdot & & & & \\ \cdot & & & & \\ \cdot & & & & \\ r_{xx}(L-1) & \cdots & & & r_{xx}(0) \end{pmatrix} \tag{27}$$

The smaller matrices are:

$$[r_{xx}(s)]_{kj} = R_{kj}(s) \tag{28}$$

and hence:

$$r_{xx}(s) = r_{xx}{}^{T}(-s) \tag{29}$$

Thus the matrix R_{xx} is actually symmetric when taken element by element. The reason for writing it in the form of equation (27) is that the standard method of solving equation (26), the Levinson recursion, (Levinson, 1949) takes advantage of this. In particular, this form is commonly called a block Toeplitz matrix. In a similar fashion the other two matrices in equation (26), F and R_{xd}, are also written in terms of smaller matrices:

$$F = \begin{pmatrix} f(0) \\ f(1) \\ \cdot \\ \cdot \\ \cdot \\ f(L-1) \end{pmatrix} \qquad R_{xd} \begin{pmatrix} r_{xd}(0) \\ r_{xd}(1) \\ \cdot \\ \cdot \\ \cdot \\ r_{xd}(L-1) \end{pmatrix} \tag{30}$$

where these smaller matrices are:

$$[f(s)]_i = f_i(s) \quad [M_{xd}(s)]_i = R_{id}(s) \tag{31}$$

Now if L is the number of points in each filter and N is the number of channels then solving the Wiener-Hopf equation in the time domain means solving N·L simultaneous linear equations:

$$F = R_{xx}^{-1} R_{xd} \tag{32}$$

However the multichannel Levinson recursion requires not $(N \cdot L)^3$ operations for the inversion of R_{xx}, but only N^3L^2 operations. This coupled with the fact that only the first column of matrices in R_{xx} need be stored in the machine makes solving the problem a practical matter.

In the frequency domain, the errors at different frequencies are independent, so they can be minimized separately. The error is written:

$$E(\omega) = |y(\omega)-d(\omega)|^2 = \sum_i f_i^*(\omega) \; S_{ij}(\omega) f_j(\omega) - \sum_i f_i^*(\omega) x_i^*(\omega) d(\omega)$$

$$- \sum_i f_i(\omega) x_i(\omega) d^*(\omega) + d^*(\omega) d(\omega) \tag{33}$$

Setting the derivations equal to zero gives:

$$\sum_i S_{ki}(\omega) \; f_i(\omega) = S_{kd}(\omega) \tag{34}$$

where the spectral matrix elements are:

$$S_{ij}(\omega) = E\{x_i^*(\omega) \; x_j(\omega)\} \tag{35}$$

This problem is a set of N simultaneous equations at each frequency. It is therefore much easier to solve the shaping problem in the frequency domain, since only LN^3 operations are required.

Prediction Filters: If the desired output is one of the inputs s points ahead in time:

$$d(t) = x_\ell(t+s) \tag{36}$$

then the filter equations reduce to:

$$\sum_\tau \sum_i R_{ki}(p-\tau)\, f_i(\tau) = R_{k\ell}(\tau+s) \tag{37}$$

and the filter is called a time prediction filter. Similarly, if the desired output is the sum of the inputs at some span ahead in time:

$$d(t) = \sum_i x_i(t+s) \tag{38}$$

the filter equations become:

$$\sum_\tau \sum_i R_{ki}(p-\tau)\, f_i(\tau) = \sum_i R_{ki}(\tau+s) \tag{39}$$

Perhaps a more interesting quantity is the prediction error. This is, for equation (36):

$$y(t) - x_\ell(t+s) \tag{40}$$

and for equation (38):

$$y(t) - \sum_i x_i(t+s) \tag{41}$$

The prediction error should be small when no signal is present in the noise and large when a signal, an unpredictable event, occurs. (Claerbout, 1964). The prediction error filter is therefore a possible detection device.

Interpolation Filters: If the desired output is another element of the array not included as an input to the filter:

$$d(t) = x_\ell(t) \tag{42}$$

the filter equations are:

$$\sum_\tau \sum_i R_{ki}(p-\tau)\, f_i(\tau) = R_{k\ell}(\tau) \tag{43}$$

This filter is called a spatial prediction or interpolation filter. The interpolation error is much like the prediction error in form being defined by:

$$y(t) - x_{\ell}(t) \tag{44}$$

When the interpolation error is small, much of the power appearing in the ℓ^{th} data channel can be represented as a multichannel convolution of the other channels. In a certain sense, then, that channel is redundant.

An important use of the interpolation filter and its associated interpolation error is in defining multiple and partial coherence functions. These are described in detail in the Appendix.

Maximum Likelihood Filters: The maximum likelihood filter, sometimes referred to as the minimum-variance unbiased filter, Levin filter, or max-like filter, (Capon and Greenfield, 1965) is the only filter presented in this report that cannot be derived directly from the Wiener-Hopf equation. It is still, however, a least-squares filter. In this section we will derive the normal equations for the maximum likelihood filter in both the time and frequency domains. It should be remembered that while both these filters are mathematically equivalent, the time-domain solution is the truly optimum version. That is, for a given filter length, the time-domain maximum likelihood filter gives optimum noise suppression while still preserving the signal. The price paid for this is in stability and the large amount of computer time needed to solve the filter equations. Nevertheless, most recent work on the maximum likelihood filter and its various uses has centered on the time-domain version. (Flinn, et al., 1966).

The mathematical trick used to specify the maximum likelihood filter equations is a summation constraint on the filter impulse response. This was first recognized by Levin,

(Kelley and Levin, 1964). If the data channels are represented by:

$$x_i(t) = n_1(t) + s(t) \tag{45}$$

where $n_1(t)$ are the noise processes and $s(t)$ is the signal, then the multichannel convolution algorithm gives:

$$y(t) = \sum_i \sum_\tau f_i(\tau)\, x_i(t-\tau) = \sum_i \sum_\tau f_i(\tau)\, n_i(t-\tau) + \sum_\tau s(t-\tau). \sum_i f_i(\tau) \tag{46}$$

Now when a summation constraint is introduced:

$$\sum_i f_i(\tau) = d(\tau) \tag{47}$$

the output can be written:

$$y(t) = \sum_i \sum_\tau f_i(\tau)\, n_i(t-\tau) + \sum_\tau d(\tau)\, s(t-\tau) \tag{48}$$

Clearly if $d(t)$ is chosen to be a spike at some time $t = p$, then the signal will be reproduced with a delay p:

$$y(t) = \sum_i \sum_\tau f_i(\tau)\, n_i(t-\tau) + s(t-p) \tag{49}$$

Among the obvious choices are putting the spike at the beginning of $d(t)$ and putting it in the middle. The resulting filters are called the realizable and symmetric maximum likelihood filters respectively. (Flinn and Claerbout, 1965). It has been observed that the symmetric maximum likelihood filter, because it is "two-sided", produces a noticeable improvement over the realizable filter. (Flinn et al., 1966). A two-sided filter is able to look ahead a bit and thus improve its performance. Since the signal part of the output signal convolved with $d(t)$, some bandpass filtering of the signal can be done without any extra effort. In this case $d(t)$ would be the impulse response of the desired filter.

The filter equations follow directly from the least squares condition modified by the constraint. The proper

mathematical technique to use here is the method of Lagrange multipliers. We want to minimize the energy:

$$E = \sum_t y(t)^2 = \sum_t \sum_i \sum_\tau f_i(\tau)\, x_i(t-\tau) \sum_j \sum_s f_j(s)\, x_j(t-s)$$

$$= \sum_i \sum_j \sum_\tau \sum_s R_{ij}(\tau - s)\, f_i(\tau)\, f_j(s) \qquad (50)$$

subject to the constraint:

$$\sum_i f_i(\tau) = d(\tau) \qquad (51)$$

The proper quantity to minimize is then:

$$E=E + \sum_\tau \lambda(\tau)\Big[\sum_i f_i(\tau)-d(\tau)\Big] = \sum_{ij\tau s} R_{ij}(\tau-s)\, f_i(\tau)\, f_j(s)$$

$$+ \sum_\tau \lambda(\tau)\Big[\sum_i f_i(\tau)-d(\tau)\Big] \qquad (52)$$

Where $\lambda(t)$ is now a new variable called the Lagrange multiplier. The derivatives with respect to $f_k(p)$ give:

$$\frac{\partial E}{\partial f_k(p)} = 0 = 2 \sum_{i\tau} R_{ki}(p-\tau) f_i(\tau) + \lambda(p) \qquad (53)$$

or, by absorbing the extra factor of -2 into $\lambda(p)$ and using the matrix notation:

$$R_{xx}F = G\lambda$$

where (54)

$$[\lambda]_p = \lambda(p)$$

Another equation is obtained by taking derivatives with respect to $\lambda(p)$:

$$\sum_i f_i(p) = d(p) \qquad (55)$$

This can be put into matrix form by defining a new matrix G composed of smaller matrices g(t):

$$G = \begin{pmatrix} g(0) \\ g(1) \\ g(2) \end{pmatrix} \qquad [g(\tau)]_{ij} = \delta_{\tau}^{j} \tag{56}$$

An example of G for N = 2 and L = 3 is:

$$G = \left(\begin{array}{ccc} 1 & 0 & 0 \\ 1 & 0 & 0 \\ - & - & - \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ - & - & - \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{array}\right) \tag{57}$$

With this definition of G, the second equation can be written:

$$G^{T}F = D \text{ or } [D]_{p} = d(p) \tag{58}$$

The two equations which specify the maximum likelihood filter are then:

$$\begin{aligned} R_{xx}F &= G\lambda \\ G^{T}F &= D \end{aligned} \tag{59}$$

These can be solved by some matrix manipulation to give:

$$F = R^{-1}G\,(G^{T}R^{-1}G)^{-1}D \tag{60}$$

In the frequency domain it can be shown that the maximum likelihood filter is defined by a similar formula. We will not actually derive it but will show that it corresponds very closely to the result obtained above. The correlation matrix becomes the spectral matrix. The matrix G is a similar summation matrix and the constraint vector reduces to a constant, usually chosen to be 1.

Thus:

$$f_i(\omega) = \sum_j S^{-1}_{ij}(\omega) \Big/ \sum_i \sum_j S^{-1}_{ij}(\omega) \qquad (61)$$

$S^{-1}(\omega)$ is taken here to be the inverse of the spectral matrix. The chief computational advantage of solving the problem in frequency is that it can be done frequency by frequency. Only the inversion of each Hermitian N-by-N spectral matrix is required.

Isotropic Processors. A further assumption usually made in least-squares filter theory which simplifies the resulting equations is that the signal and noise are uncorrelated or incoherent. Now, if the inputs into equation (34) are taken to be:

$$x_i(\omega) = N_i(\omega) + Q_i(\omega) \qquad (62)$$

this assumption reduces equation (34) to:

$$\sum_i [N_{ki}(\omega) + Q_{ki}(\omega)] f_i(\omega) = S_{kd}(\omega) \qquad (63)$$

The cross term cancels out. Furthermore if the desired output is chosen to be some sort of signal the equation is:

$$\sum_i [N_{ki}(\omega) + Q_{ki}(\omega)] f_i(\omega) = Q_{kd}(\omega) \qquad (64)$$

where N is the spectral matrix of the noise and Q is the spectral matrix of the signal. It has been shown by Burg, (Burg, 1964; Spieker, et al., 1961) that good filters can be designed from equation (64) using an analytical signal spectral matrix and either an analytical or a measured-noise spectral matrix. These filters we will call the theoretical and measured noise isotropic processors.

For analytical models, we will calculate what the spectral matrix elements are for given velocity distributions. In all the following cases the velocity distribution is assumed to be independent of frequency. The first model is the point

model. Data traveling with a given vector velocity $\bar{v}$ has spectral matrix elements of the form:

$$S_{ij}(f) = P(f) \exp\left[\frac{2\pi i f \bar{v}}{|\bar{v}|^2} \cdot (\bar{x}_i - \bar{x}_j)\right] \tag{65}$$

Here P(f) is the power spectrum which is assumed to be the same for all channels. The ring model for data with a given scalar velocity v is obtained by integrating equation (65):

$$\frac{1}{2\pi k} \int_0^{2\pi} \exp[2\pi i |\bar{k}| \, |\bar{x}_i - \bar{x}_j| \cos\theta] \, k d\theta = J_o(2\pi |\bar{k}| \, |\bar{x}_i - \bar{x}_j|) \tag{66}$$

The ring model therefore has spectra of the form:

$$S_{ij}(f) = P(f) \, J_o \left(\frac{2\pi f}{v} |\bar{x}_i - \bar{x}_j|\right) \tag{67}$$

Two other models can be derived by integrating equation (65) over certain regions in K space. These are the disc model, $v \geq v_s$, and the annular ring model, $v_L \leq v \leq v_H$. The spectra for them are:

$$S_{ij}(f) = \frac{P(f) v_s}{\pi f |\bar{x}_i - \bar{x}_j|} J_1 \left(\frac{2\pi f}{v_s} |\bar{x}_i - \bar{x}_j|\right) \tag{68}$$

$$S_{ij}(f) = \frac{P(f)}{\pi f^2 \left(\frac{1}{v_H^2} - \frac{1}{v_L^2}\right)} |\bar{x}_i - \bar{x}_j| \left[\frac{f}{v_H} J_1 \left(\frac{2\pi f}{v_H} \bar{x}_i - \bar{x}_j\right)\right. \tag{69}$$

$$\left. - \frac{f}{v_L} J_1 \left(\frac{2\pi f}{v_L} |\bar{x}_i - \bar{x}_j|\right)\right]$$

There are two other spectral matrices which are useful to define. These are the infinite velocity and incoherent matrices. Their elements are:

$$S_{ij}(f) = P(f) \tag{70}$$

$$S_{ij}(f) = P(f)\ \delta_j^i \tag{71}$$

The scheme therefore is to insert combinations of these models into equation (64). As a simplification, the desired output is chosen to be the signal which would be seen at the origin of the array. $\bar{x}_d = 0$. A parameter is also inserted to allow varying the signal to noise ratio of the filters. With these assumptions equation 64 is:

$$\sum_i [N_{ki}(f) + c\ Q_{ki}(f)]\ f_i(f) = Q_{ko}(f). \tag{72}$$

The theoretical isotropic processor uses an annular ring model for the noise and either a disc or an infinite velocity signal model. P(f) is chosen to be 1 in this case. For the measured noise isotropic processor, an actual spectral matrix computed from a sample of noise is used with the same choice of signal models. However here, P(f) is chosen to be a certain noise power spectrum.

One further modification of the signal models is necessary to make the resulting filter response functions stable. In actual practice the relative magnitudes of the input data channels, or gain corrections, are not accurately known. The least-squares equations can take advantage of this fact and give solutions which do not in general reduce the noise level. Another way to see this is that the signal matrix Q is nearly singular. It specifies the signal model too accurately. Consequently some relaxation of the signal model is necessary to stabilize the computation. The most natural way to do this is to add a few percent white noise to the model. A good figure is three percent, which can be shown to allow an even distribution of gain corrections between .7 and 1.3. Our programs are designed to do this.

RESULTS

The results of this study are presented in Table 3 and

in Figures 3 through 10. The first three figures show characteristics of the array and the data; the rest of the figures show multichannel filter outputs and evaluations. Table 3 lists the actual S/N improvement obtained by each method. Viewed together, these results represent a large amount of information. This section will explain each result and also discuss the various comparisons which can be made.

The array response function, contour-plotted in Figure 3, demonstrates the basic capability of the array. To some degree it is reflected in all of multichannel filter f-k response plots that follow. This response has a reasonably small central lobe but also six prominent side lobes. The number of side lobes reflects the somewhat triangular nature of the array and their amplitude is a direct result of only using ten sensors. The sharp central lobe comes from using well-spaced sensors. Adding more sensors would reduce the side lobes by a gradual averaging process but would not appreciably increase the wavenumber resolution of the array. Generally speaking, the expected result of beamforming this array would be good signal enhancement at high frequencies but mediocre noise rejection; since noise will leak in though the side lobes.

Plots of the actual data are shown in Figures 4 and 5. Figure 4 is the raw data and Figure 5 is the same data bandpass filtered. The incident P wave used as the signal can be seen towards the end of both seismograms. The amount of noise in front of the signal in both figures is not representative of all there was on the tape. In actuality, approximately 200 seconds (4096 points at 20 samples per second) of noise were available to the programs.

The noise f-k spectra shown for four frequencies in Figure 8 were computed using all 200 seconds of noise. The raw spectra in the spectral matrix were smoothed and decimated five times from 4097 to 129 points with our triangular

convolution and decimation procedure. This gave 32 degrees of freedom to each frequency estimate and a resolution of 0.078 cps.

The most salient feature of those four plots is the presence of high-velocity seismic noise. Any of the multichannel filters computed from a measured noise sample should experience difficulty in enhancing a high velocity signal and rejecting high velocity noise. A general characteristic of these noise spectra is the influence of the array response. This effect is most clearly seen in the 0.39 cps plot where the area enclosed by the -3 db contour is almost a third of the whole plot. At high frequencies the effect of the array response is smaller but above the 1.25 cps the noise field is almost flat and therefore uninteresting. The meaning of all this is evident: this array is just not suitable for investigating low-frequency noise.

Figures 6 and 7 show the output of various multichannel filters as well as the straight sums for un-prefiltered and prefiltered data respectfully. The fitting interval for the maximum-likelihood filter was 150 seconds, ending just short of the signal arrival time. The fitting interval for the measured-noise isotropic processors was approximately 100 seconds (2048 points at 20 samples per second), ending before the origin time of the plot. The theoretical isotropic processor, of course, is not computed from any actual data.

The maximum likelihood filter was thirty points long sampled at 10 samples per second giving it an effective length of three seconds. It required 31 minutes of CDC 1604B time to calculate, 16 minutes for the correlation matrix and 15 minutes for the filter. The theoretical and measured-noise isotropic processors were 255 points long, sampled at 20 samples per second, giving them effective length of almost thirteen seconds. The measured-noise isotropic processors required 35 minutes of computer time to calculate, 20 minutes

for the spectral matrix and 15 minutes for the filter. The theoretical isotropic processors required 12 minutes of computer time to calculate.

Table 3 lists the actual improvement in S/N ratio over the un-prefiltered straight sum for the various multichannel filters. Simply bandpass filtering the sum gets 15 db improvement in S/N ratio at practically no expense of computer time, so the real comparison is with this reference level. The amplitude of the noise was computed for the measured noise multichannel filter outputs over an interval of 200 seconds, including the fitting interval and 100 seconds outside the fitting interval. The maximum-likelihood noise amplitude was also computed over an interval of 200 seconds, but this time including the fitting interval and 50 seconds outside it. Therefore, these noise levels represent an average of the rms noise amplitude inside fitting interval and outside it. This problem does not, of course, arise with the straight sums and theoretical isotropic processor outputs.

We can make several comparisons in this table. The first concerns the amount of frequency filtering that is done. The illusory gains from the maximum-likelihood filter operating on un-prefiltered noise come about because the least-squares equation does not constrain the filter to do otherwise. Many degrees of freedom in the filter weights are misused doing what could be done with just elementary processing, i.e., bandpass filtering. Apparently all but 2 db of the original 15 db can be accounted for by bandpass filtering. On the other hand, the design equation for the measured-noise multichannel filter specifies as a desired signal spectrum one of the noise spectra. Therefore, some frequency filtering is expected. The design equation for the theoretical multichannel filter assumes a flat spectrum for the desired signal output. However, the filter which

satisfies this equation is only the optimum filter in the least-squares sense, so we should expect some but not much frequency filtering. This is demonstrated in the table. In summary then the maximum likelihood multichannel filter is the worst regarding frequency filtering and the theoretical multichannel filter is the best.

Secondly, in both cases, the theoretical and the measured noise filters designed with a disc signal model ($V \geq$ 10 km/sec) performed better than the infinite-velocity filters. This fact is true in general and is a direct result of not beamsteering the array. However, other results discussed below suggest that this gain occurs only in a frequency band higher than the signal band. For the prefiltered theoretical multichannel filter, designed on noise where this energy has been removed, this difference only amounts to 0.3 db.

Finally, as regards isotropic processor multichannel filters, the theoretical processor is always better than measured-noise processor. This is a strange and completely unexpected result, probably arising from the fact that the noise seems to overlap the signal model in k space, as Figure 8 shows. For a measured-noise multichannel filter this represents a contradiction of specification in the least-squares design equations which could lend again to a misuse of the degrees of freedom available in the filter weights. The theoretical multichannel filter is not computed from actual data and hence is not affected.

Figure 9 shows, for two frequencies, the f-k transfer functions of each multichannel filter. In this figure alone the direction of the k vector is opposite to that of our original definition and the definition in the other figures. Here the k vector points in the direction toward which an equivalent wave front would be moving. These plots show clearly how for both isotropic processors the infinite-velocity signal model provides a sharper signal band at 1.25

cps. The passband of the theoretical multichannel filters, 2 to 4 km/sec., is also evident in these plots. Although comparison is difficult because of the difference in the direction of the k vector, it is possible to see that some of the noise peaks in Figure 8 are nulled by the measured-noise multichannel filter f-k transfer function. Unfortunately not all the noise is reduced. It is also possible to see small differences in the k-space passbands due to changes from un-prefiltered to prefiltered noise in the pertinent multichannel filter fitting intervals. Perhaps nulling noise by a strong bandpass filter actually destroys some of its spatial information. In passing, we note that another deficiency of the maximum-likelihood filter can be seen here, for more than any other multichannel filter the maximum-likelihood multichannel filter produces a complicated and generally strange k-space pattern. It may be that many degrees of freedom are lost in generating this unnecessarily sharp response.

Figure 10 shows plots of the performance factor of each multichannel filter (in terms of S/N ratio improvement) to a perfectly correlated signal in the presence of completely uncorrelated noise. The performance factor is defined by:

$$P(\omega) = \frac{\left|\sum_i f_i(\omega)\right|^2}{\sum_i \left|f_i(\omega)\right|^2} \qquad (73)$$

These plots confirm the earlier conjecture that the infinite-velocity signal model isotropic processors work best on noise higher than 1 cps. This is most dramatic in the case of the theoretical multichannel filter where the performance tends asymtotically to 1. Here again is illustrated the general strangeness of the maximum-likelihood multichannel filter; whose performance factors stand out as unlike the others.

One particular difficulty of the maximum-likelihood multichannel filter is its poor performance at low frequencies; which is most likely due to its relatively short impulse response (three seconds as compared to thirteen seconds for the other filters).

CONCLUSIONS

The conclusions listed below are based upon results obtained from processing only one event. In view of this they should be accepted with due caution.

1. Designing multichannel filters on un-prefiltered noise gives an illusory S/N ratio improvement which is not attained with similar multichannel filters designed on prefiltered noise.

2. Isotropic processors can be effectively applied to un-beamsteered arrays, provided a finite velocity signal model is used.

3. Measured-noise isotropic processors are not effective in improving S/N ratio when the noise has large high-velocity components. In this case the theoretical multichannel filter is the superior processor.

4. With the exception of the maximum-likelihood filter evaluated for performance within its fitting interval, only the theoretical multichannel filter showed a gain over the prefiltered straight sum. That gain was slightly more than 1 db, which is not regarded as significant.

5. One drawback of the maximum-likelihood filter is its unnecessarily complex response in k-space. It does not appear that modifying the constraint will cure this.

APPENDIX

Multiple and Partial Coherence Functions. The multiple coherence function of a data channel is a measure of how much power on that channel can be interpolated from a set of other channels. (Bendat and Piersol, 1966). Imagine an array configuration of four sensors at the corners of a square:

$$\begin{matrix} A & x_2 \\ x_1 & x_3 \end{matrix}$$

The interpolation filter equations to interpolate channel A from channels x_1, x_2, and x_3 are:

$$S_{xx}(\omega)\, F(\omega) = S_{xA}(\omega) \tag{74}$$

or written out:

$$\begin{pmatrix} S_{11} & S_{12} & S_{13} \\ S_{21} & S_{22} & S_{23} \\ S_{31} & S_{32} & S_{33} \end{pmatrix} \begin{pmatrix} F_1 \\ F_2 \\ F_3 \end{pmatrix} = \begin{pmatrix} S_{1A} \\ S_{2A} \\ S_{3A} \end{pmatrix} \tag{75}$$

where the ω dependence of all terms is now understood. The interpolated output $\hat{A}$ is then:

$$\hat{A} = x_1F_1 + x_2F_2 + x_3F_3 \tag{76}$$

from the multichannel convolution algorithm. If the filter F is able to completely interpolate A, in the least-squares sense, then the error:

$$A - \hat{A} \tag{77}$$

will be incoherent with A. That is:

$$E\,\{A(\omega)[A(\omega) - \hat{A}(\omega)]^*\} = 0 \tag{78}$$

which is:

$$S_{AA} = S_{A\hat{A}}$$

Now because of the convolution algorithm this can be written:

$$S_{AA} = S_{A1}F_1 + S_{A2}F_2 + S_{A3}F_3 \qquad (79)$$

Of course this is still only if the filter can account for all of the predictable power in A. A set of equations that combines both this constraint and the filter equations themselves is:

$$\left(\begin{array}{cccc} S_{AA} & S_{A1} & S_{A2} & S_{A3} \\ \hline S_{1A} & S_{11} & S_{12} & S_{13} \\ S_{2A} & S_{21} & S_{22} & S_{23} \\ S_{3A} & S_{31} & S_{32} & S_{33} \end{array}\right) \left(\begin{array}{c} 1 \\ \hline -F_1 \\ -F_2 \\ -F_3 \end{array}\right) = 0 \qquad (80)$$

Expansion of equation (80) along the indicated partition gives both equations (75) and (79). In matrix form this equation is:

$$S' F' = 0 \qquad (81)$$

where S' is now defined to be the augmented spectral matrix. For a non-trivial solution, the augmented spectral matrix must be singular. An expression for its determinant is:

$$S_{A|x} \equiv S_{AA} - (S_{A1}\ S_{A2}\ S_{A3})S^{-1} \begin{pmatrix} S_{1A} \\ S_{2A} \\ S_{3A} \end{pmatrix} = \frac{\det S'}{\det S} \qquad (82)$$

The multiple coherence function is a relative measure of this determinant or equivalently, of the solubility of these

equations:

$$\gamma^2_{A|x} \equiv 1 - \frac{S_{A|x}}{S_{AA}} = 1 - \frac{\det S'}{(\det S)S_{AA}} \qquad (83)$$

When the condition (81) and hence the conditions (75) and (79), is true, det $S' = 0$ and the multiple coherence is 1. When less and less power on channel A can be interpolated from the other channels, the multiple coherence tends toward zero.

The partial coherence function (Bendat and Piersol, 1966) is the oridnary coherence between two interpolation errors. Imagine an array configuration with elements the corners of a square.

A x_1

B x_2

The filter equations to interpolate A and B from x_1 and x_2 are:

$$\begin{pmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{pmatrix} \begin{pmatrix} F_{1A} & F_{1B} \\ F_{2A} & F_{2B} \end{pmatrix} = \begin{pmatrix} S_{1A} & S_{1B} \\ S_{2A} & S_{2B} \end{pmatrix} \qquad (84)$$

which are in matrix notation:

$$S \; F = G \qquad (85)$$

The interpolated outputs are now:

$$\hat{A} = F_{1A} x_1 + F_{2A} x_2$$

$$\hat{B} = F_{1B} x_1 + F_{2B} x_2 \qquad (86)$$

or:

$$\begin{pmatrix} \hat{A} \\ \hat{B} \end{pmatrix} = F^T \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \qquad (87)$$

The two errors in interpolation are:

$$E = \begin{pmatrix} E_A \\ E_B \end{pmatrix} = \begin{pmatrix} \hat{A} \\ \hat{B} \end{pmatrix} - \begin{pmatrix} A \\ B \end{pmatrix} \tag{88}$$

The spectral matrix of these two errors, which is frequently called the conditioned spectral matrix, is:

$$S_c^T = EE^+ = \left[\begin{pmatrix} \hat{A} \\ \hat{B} \end{pmatrix} - \begin{pmatrix} A \\ B \end{pmatrix} \right] \quad \left[(\hat{A}^* \ \hat{B}^*) - (A^* \ B^*) \right] \tag{89}$$

where of course expected values are to be taken. The transpose occurs on the left hand side of equation (89) because this definition of spectra is the complex conjugate of our previous definition, equation (19).

$$\begin{aligned} S_c^T &= S_{AB}^T + F^T S^T F^{T^+} - G^T F^{T^+} - F^T G^{+^T} \\ S_c &= S_{AB} + F^+ SF - F^+ G - G^+ F \\ &= S_{AB} + G^+ S^{-1} G - G^+ S^{-1} G - G^+ S^{-1} G \\ S_c &= S_{AB} - G^+ S^{-1} G \end{aligned} \tag{90}$$

The partial coherence between channels A and B is the ordinary coherence of this 2-by-2 matrix:

$$\gamma^2_{A,B|x_1,x_2} = \frac{S_c(1,2)\, S_c(2,1)}{S_c(1,1)\, S_c(2,2)} \tag{91}$$

REFERENCES

Bendat, J.S., and Piersol, A.G., 1966, Measurement and Analysis of Random Data, John Wiley and Sons, New York.

Blackman, R.B., and Tukey, J.W., 1959, Measurement of Power Spectra, Dover Publications, Inc., New York.

Burg, J.P., 1964, Three Dimensional Filtering with An Array of Seismometers: Geophysics, Vol. 29, No. 5, pp 693-713.

Capon, J., and Greenfield, R.J., 1965, Asympototically Optimum Multidimensional Filtering for Sampled-Data Processing of Seismic Arrays: Tech. Note 1965-67, MIT Lincoln Laboratory, Lexington, Mass.

Claerbout, J.F., 1964, Detection of P-Waves from Weak Sources at Great Distances: Geophysics, Vol. 29, pp. 197-210.

Cooley, J.W., and Tukey, J.W., 1965, An Algorithm for the Machine Calculation of Complex Fourier Series: Math. of Comp., Vol. 19, pp. 297-301.

Cooley, J.W., et al, 1967, The Fast Fourier Transform Algorithm Adaptation and Its Application: Research Paper RC 1743, IBM Research Center, Yorktown Heights, New York.

Flinn, E.A., et al., 1966, Two Examples of Maximum-Likelihood Filtering of LASA Seismograms: Seismic Data Laboratory Report No. 148, Earth Sciences Division, Teledyne, Inc., Alexandria, Virginia.

Flinn, E.A., and Claerbout, J.F., 1965, Some Topics in Digital Filtering Theory and Application: Research Dept. Memo., Earth Sciences Division, Teledyne, Inc., Alexandria, Virginia.

Kelley, E.J., and Levin, M.J., 1964, Signal Parameter Estimation for Seismometer Arrays: Tech. Report 339, MIT Lincoln Laboratory, Lexington, Mass.

Levinson, N., 1949, The Wiener RMS Error Criterion in Filter Design and Prediction: Appendix B in: Extrapolation, Interpolation, and Smoothing of Stationary Time Series, by Norbert Wiener, MIT Press, Cambridge.

McCowan, D.W., 1966, Finite Fourier Transform Theory and Its Application to the Computation of Correlations. Convolutions, and Spectra: SDL Report No. 168 (Revised) 1967, Earth Sciences Division, Teledyne, Inc., Alexandria, Virginia.

Spieker, V.J., et al., 1961, Seismometer Array & Data Processing System: Final Report. Phase I, AFTAC Project No. VT/077, Contract AF 33(600)-41840, Texas Inst., Inc., Dallas, Texas.

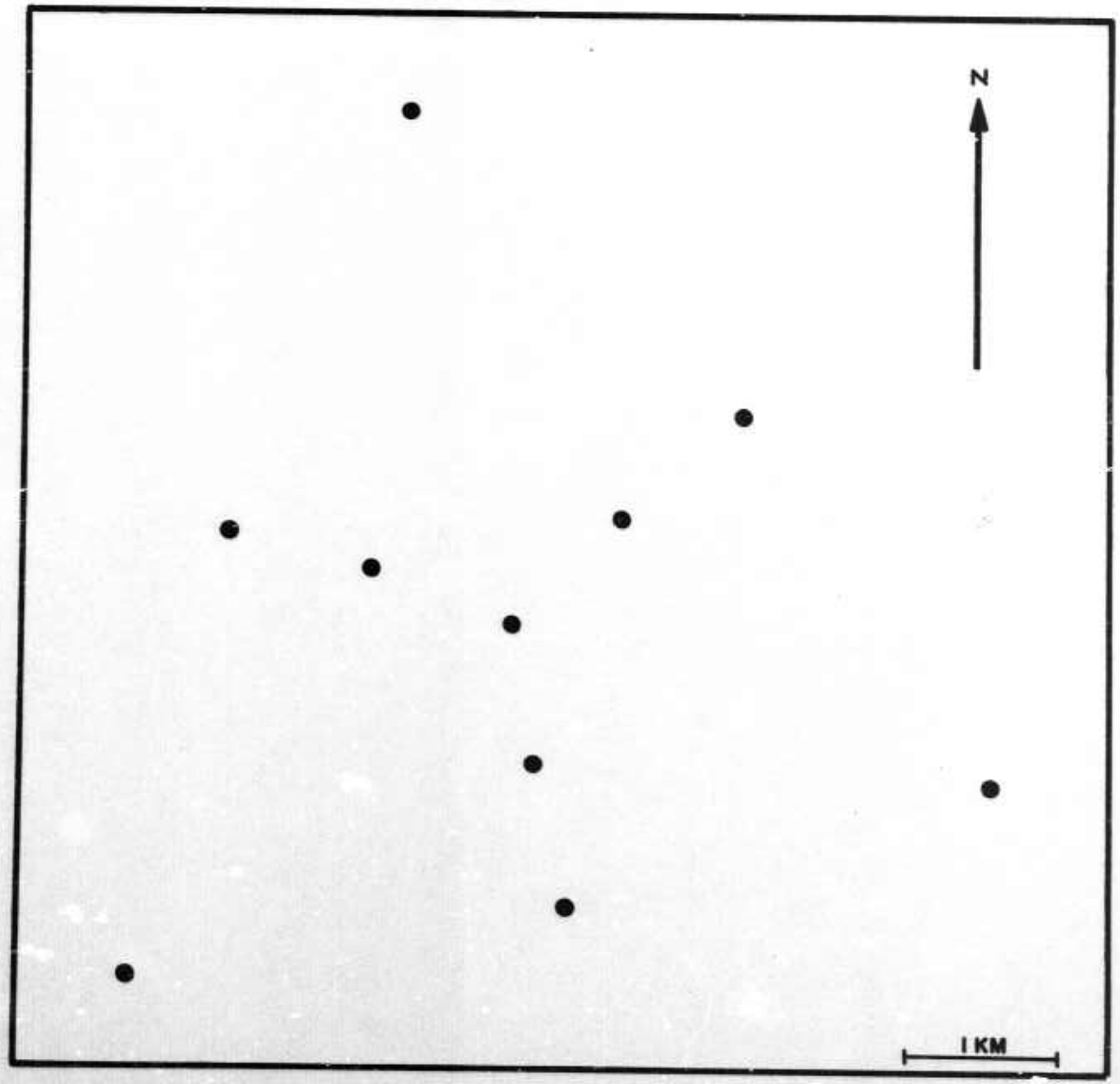


Figure 1. Array Map

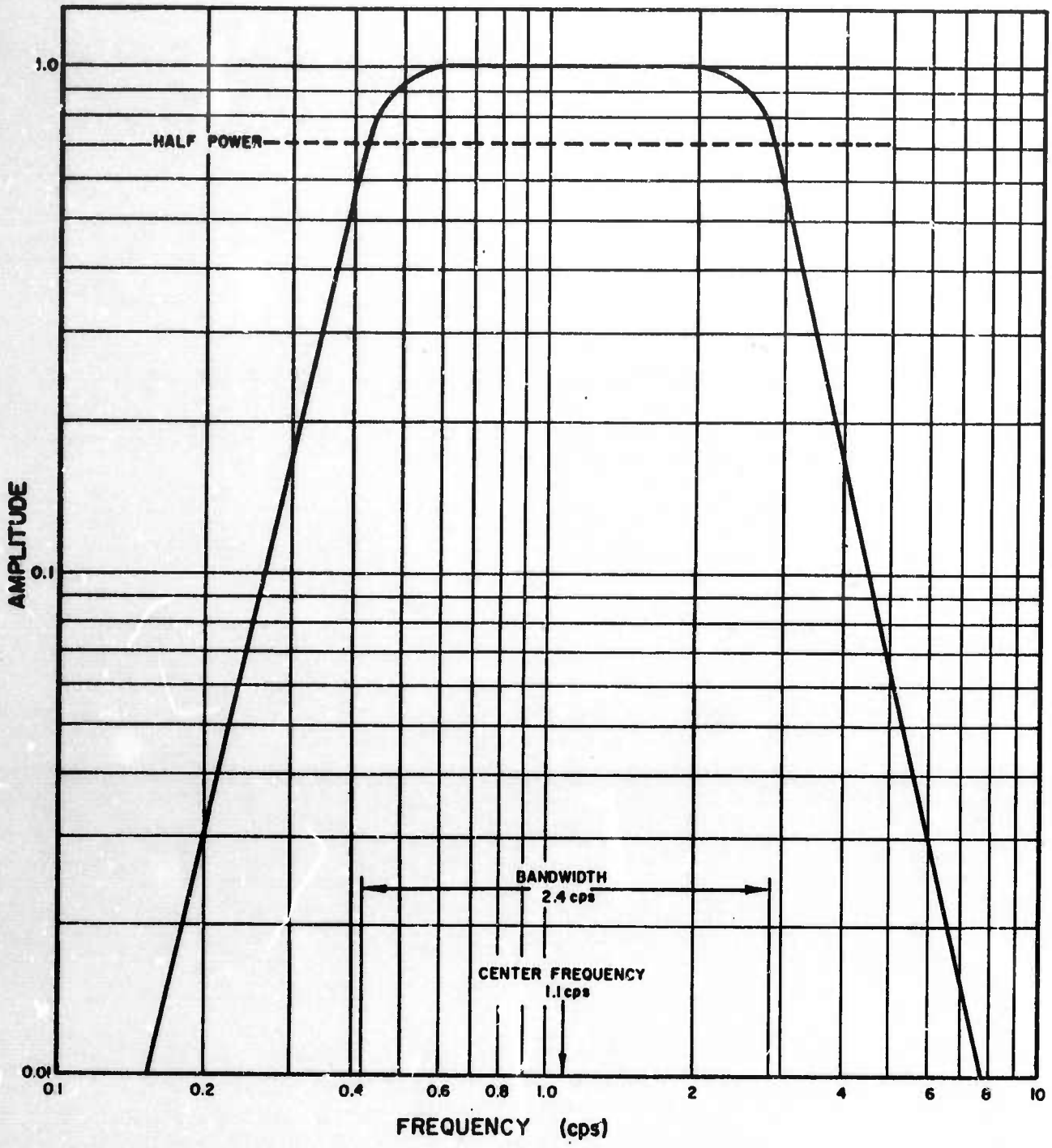


Figure 2. Bandpass Filter Used in Data Preparation

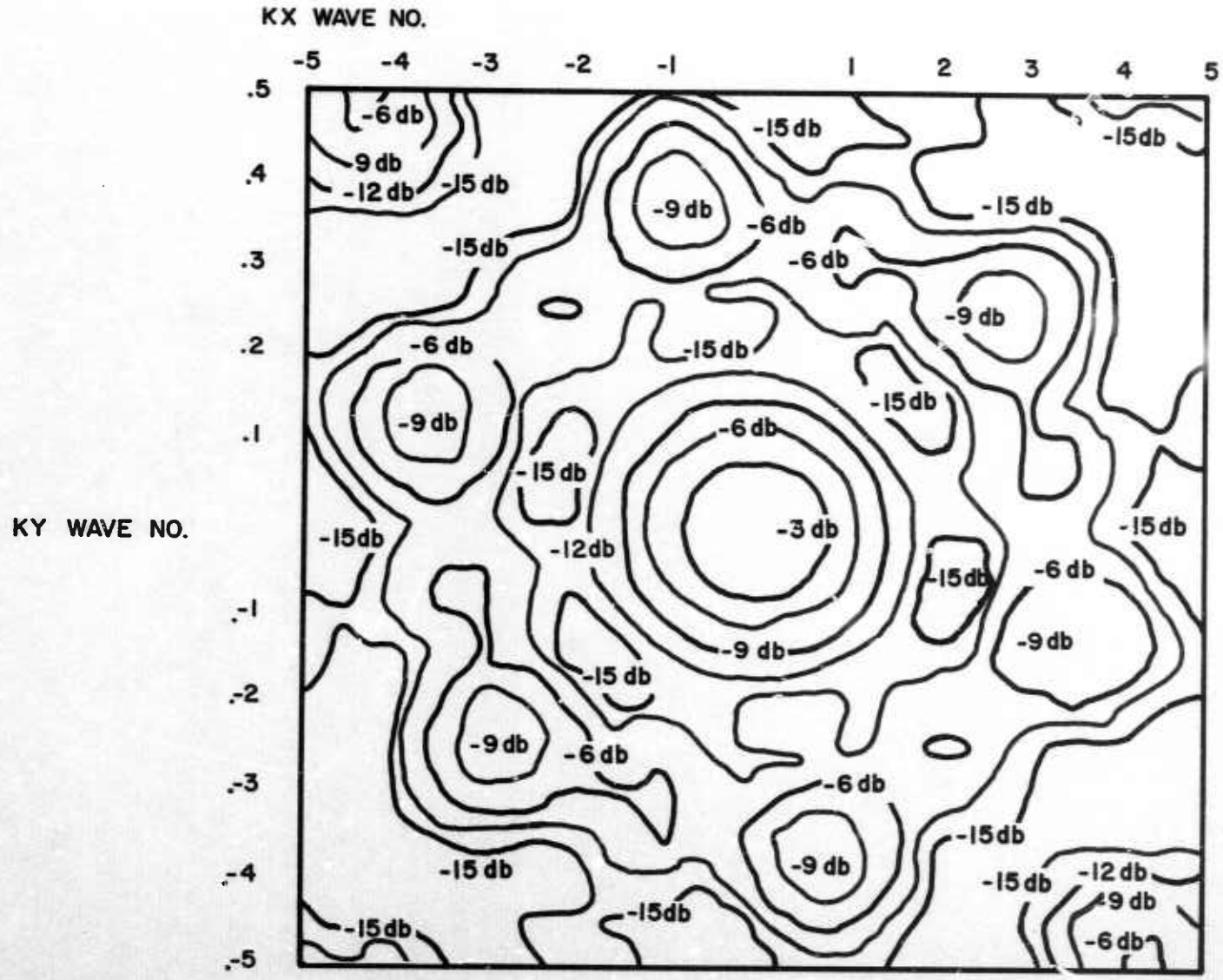


Figure 3. Array Impulse Response

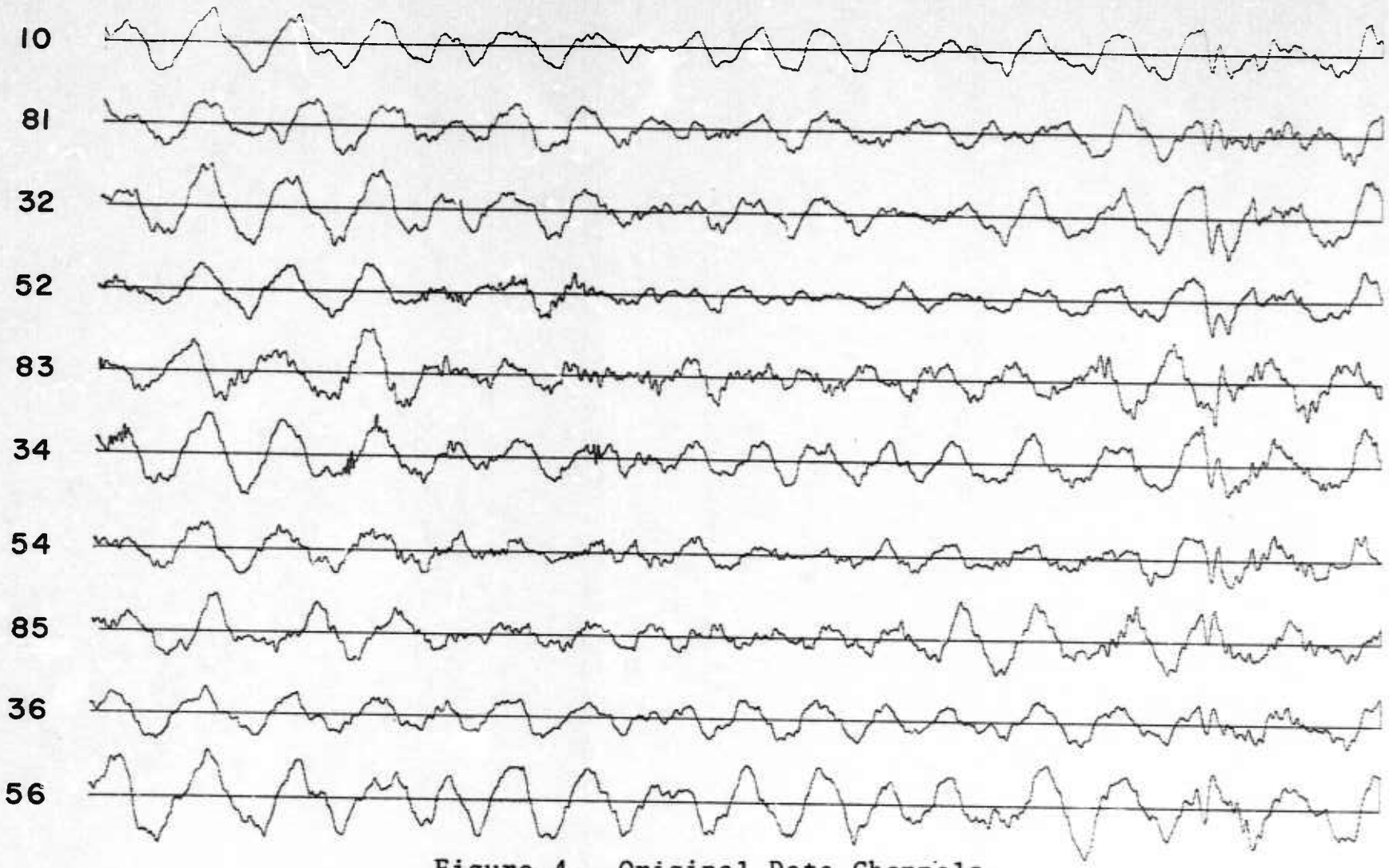


Figure 4. Original Data Channels

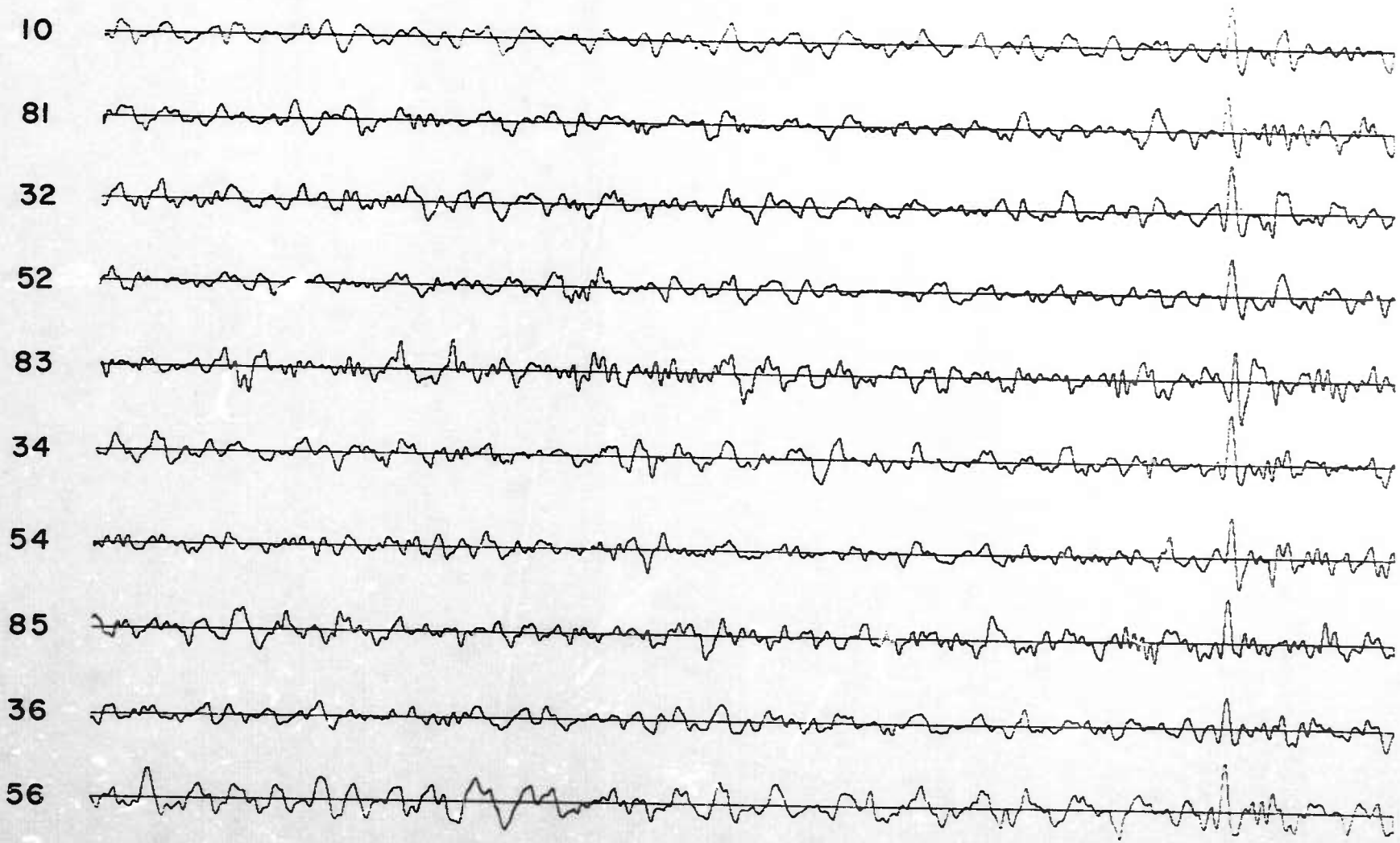


Figure 5. Prefiltered Data Channels

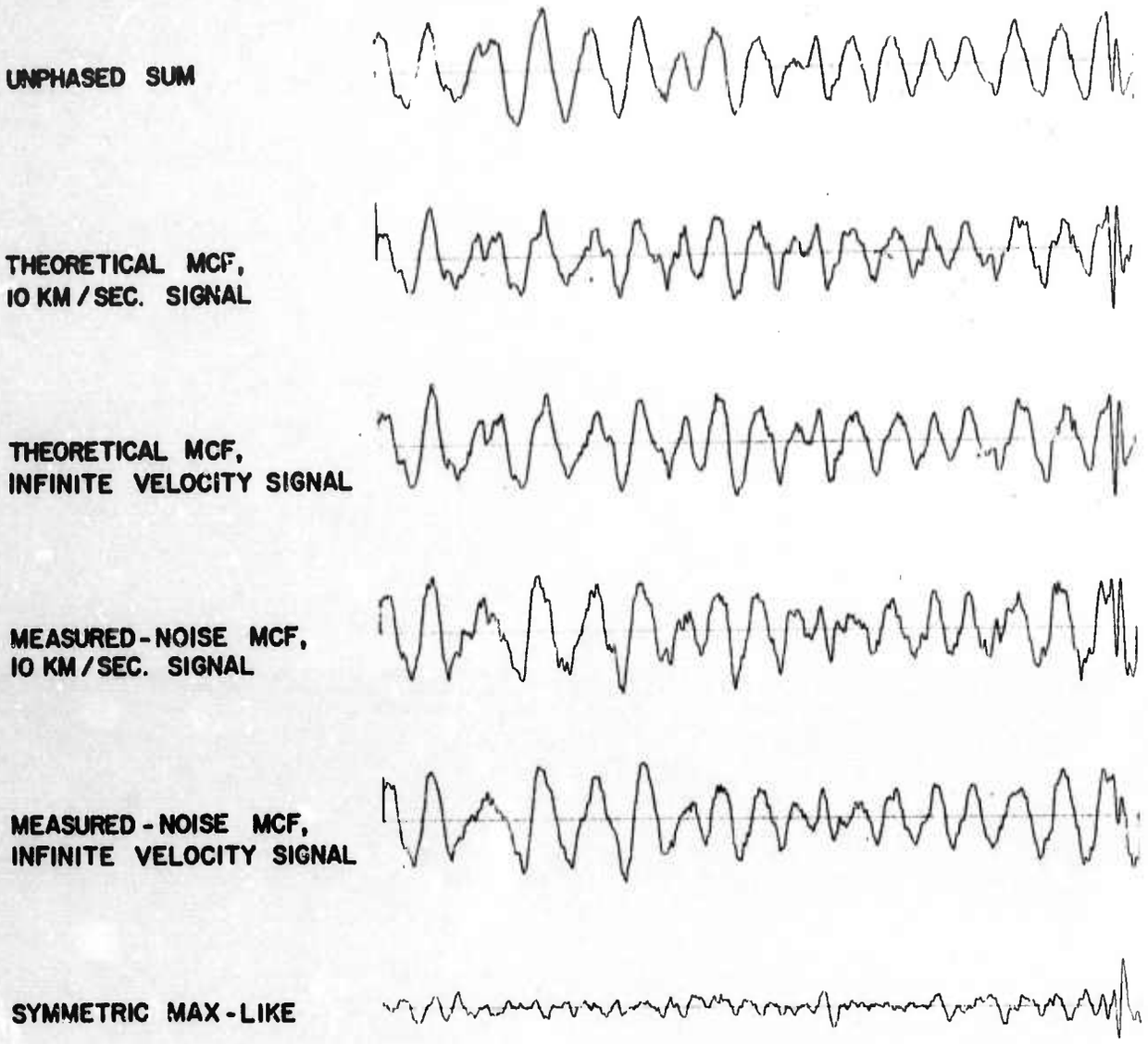


Figure 6 - Filters Designed From and Applied to Un-prefiltered data.

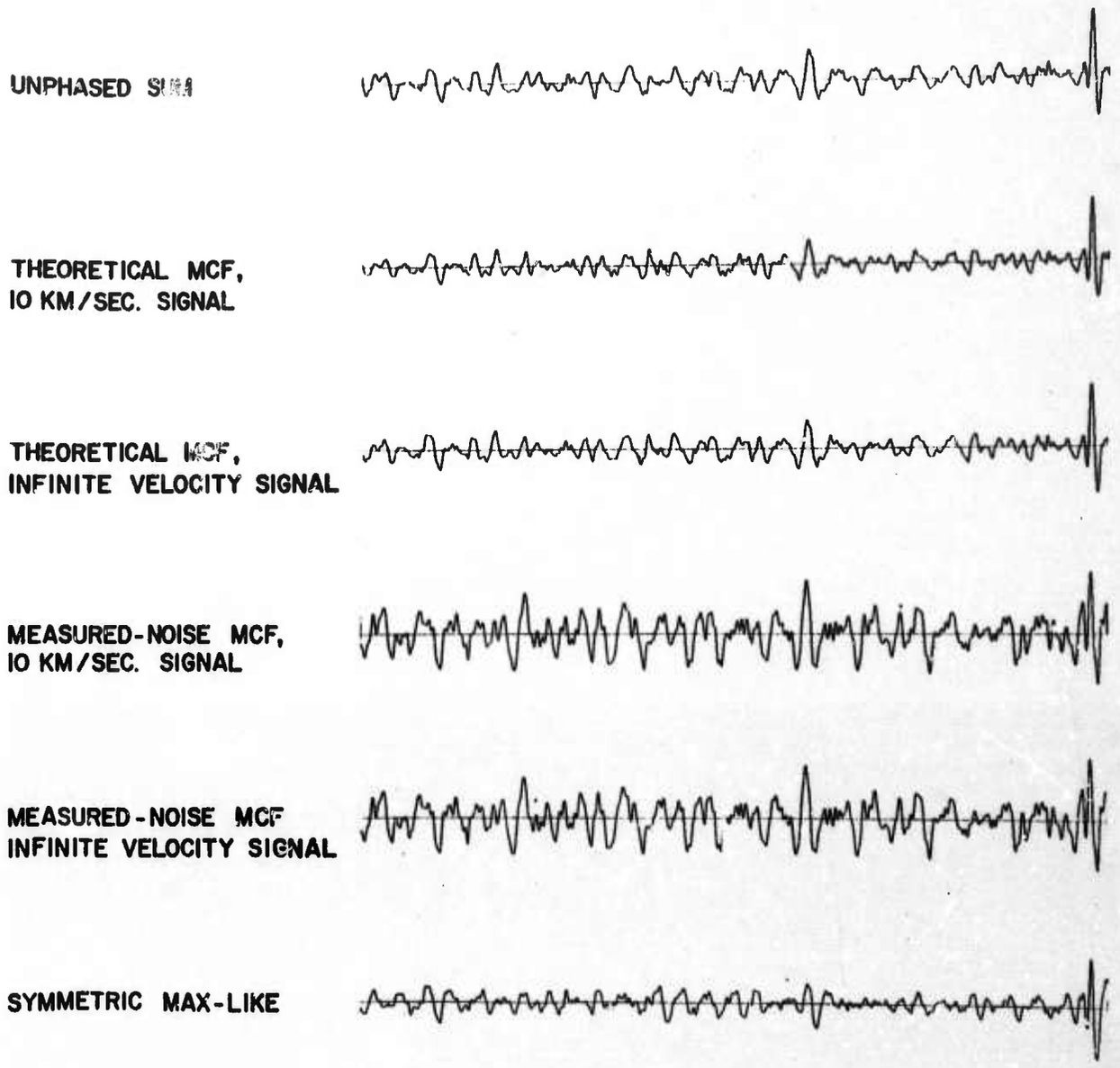


Figure 7 - Filters Designed From and Applied to Prefiltered data.

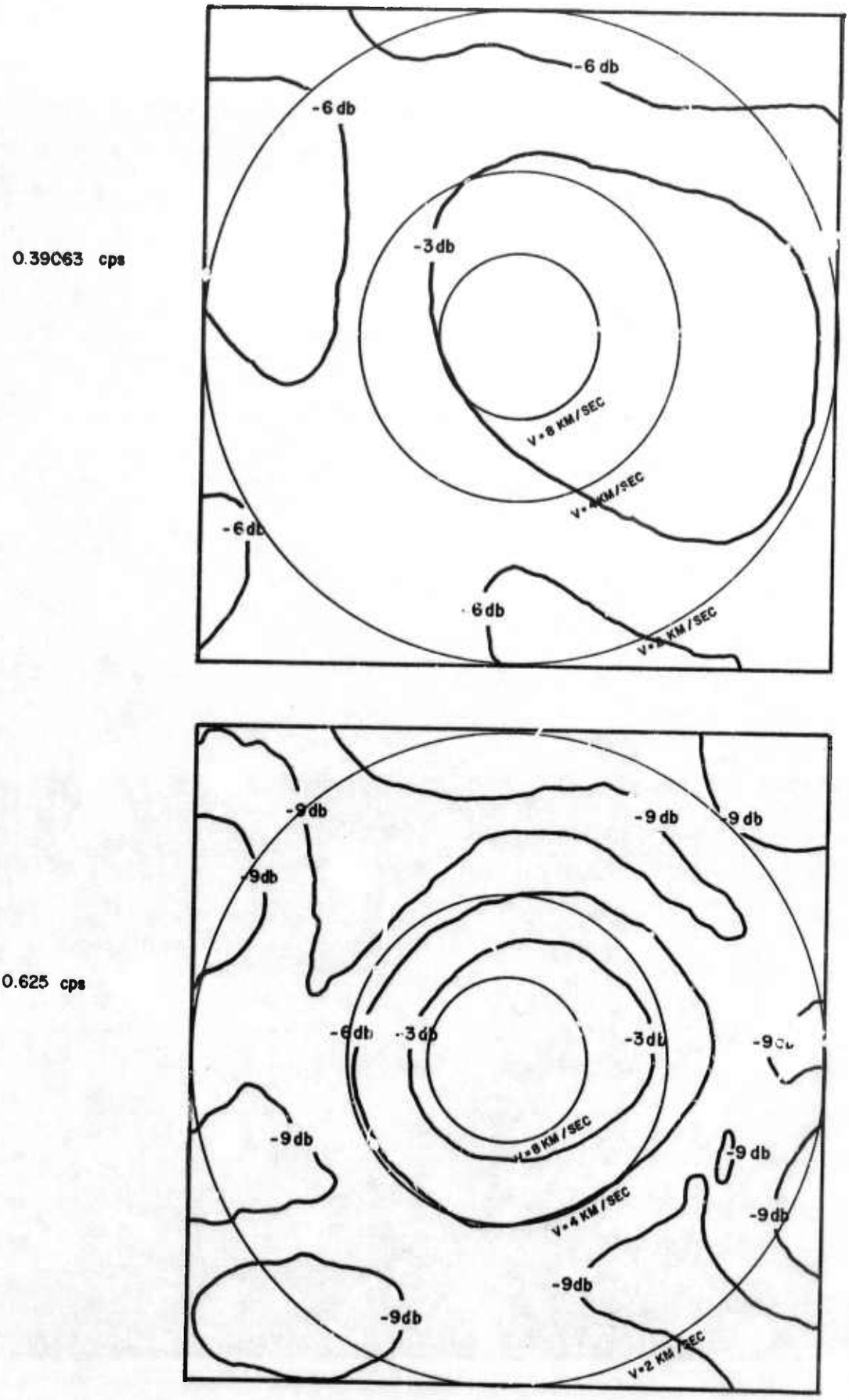
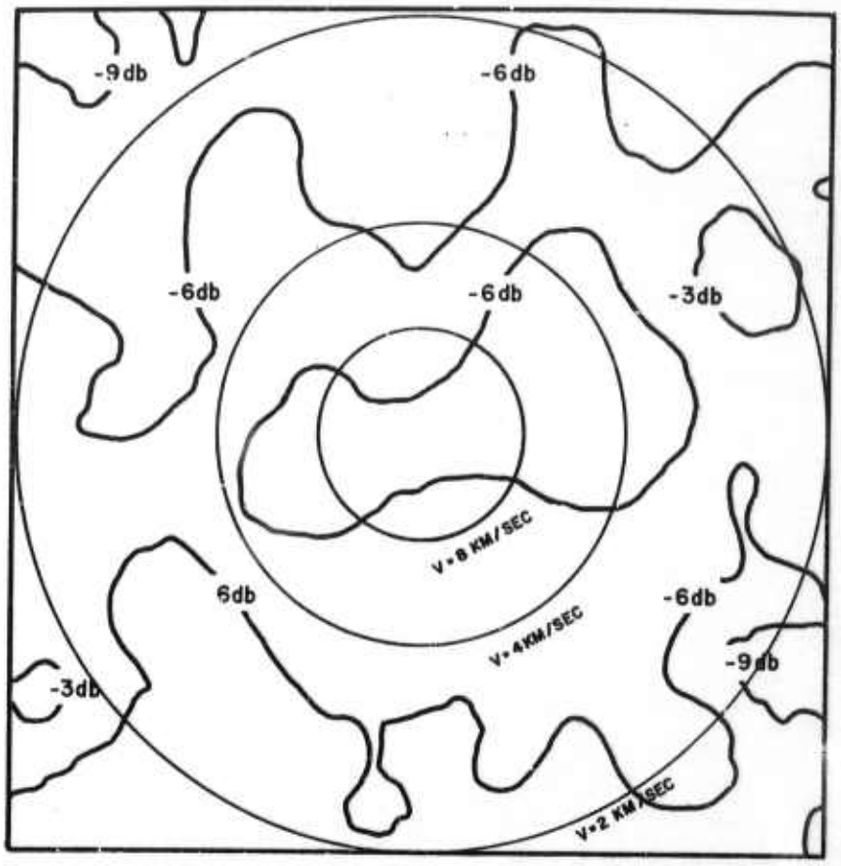
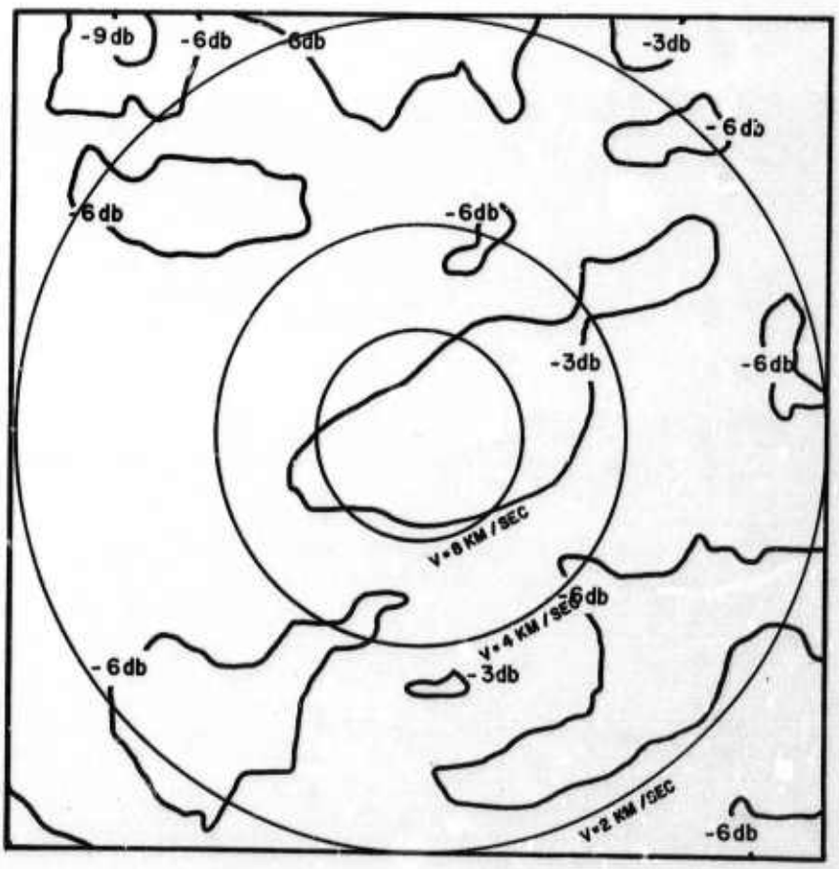


Figure 8a. FK Spectra of Data



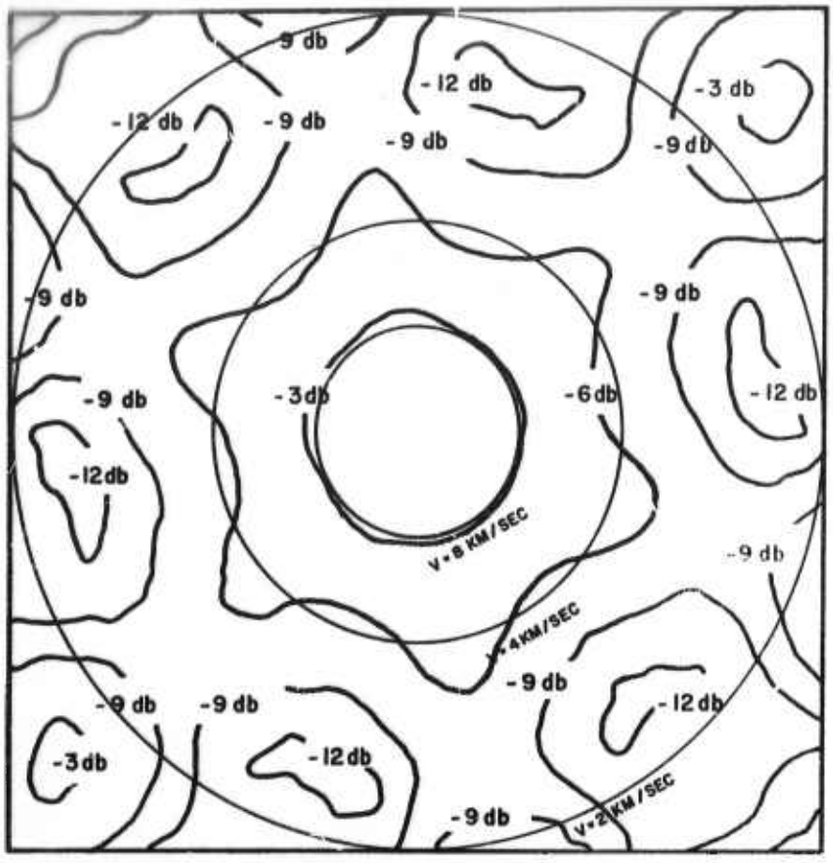
.9375 cps



1.25 cps

Figure 8b. FK Spectra of Data

0.625 cps



1.25 cps

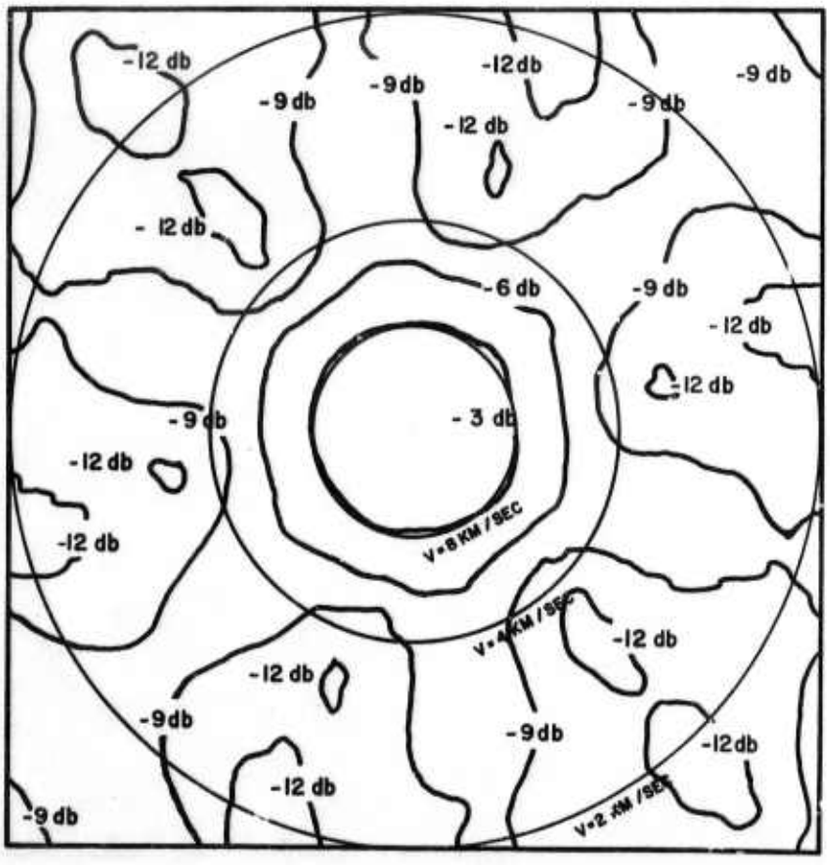
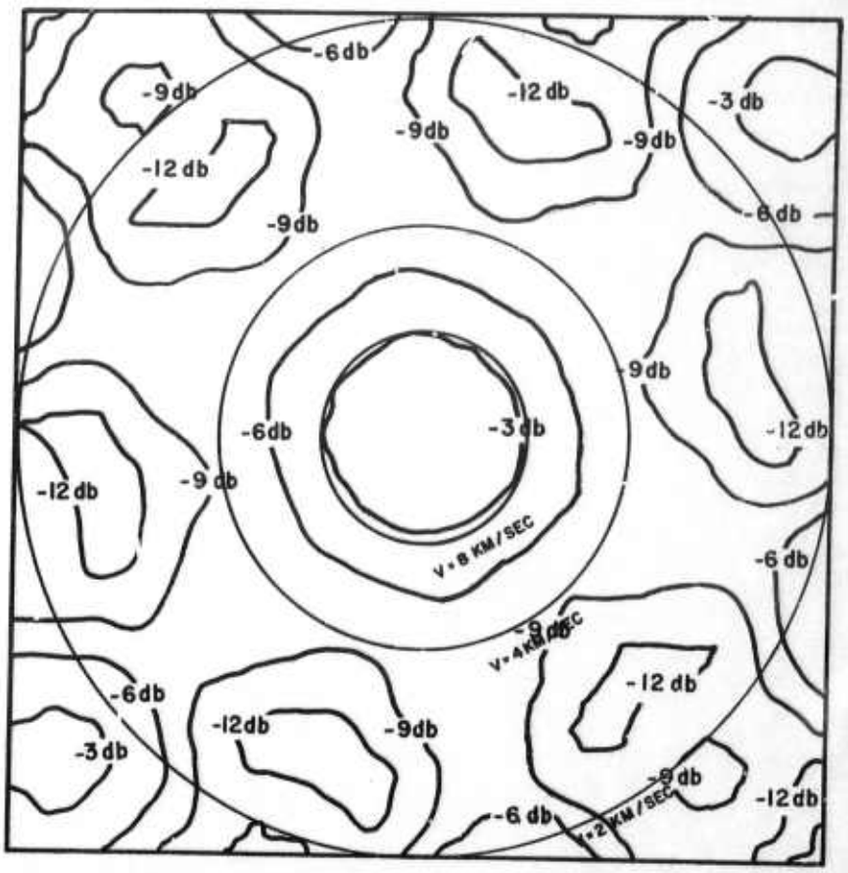


Figure 9a. FK Response of the Theoretical Processor.
10 km/sec Signal Model

0.625 cps



1.25 cps

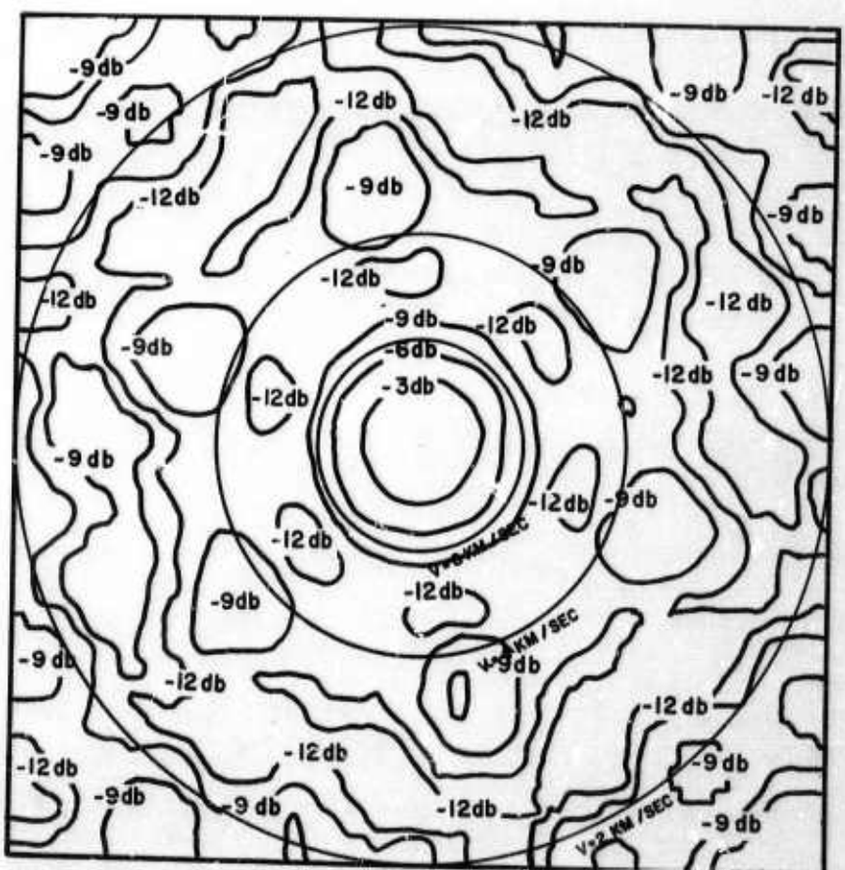
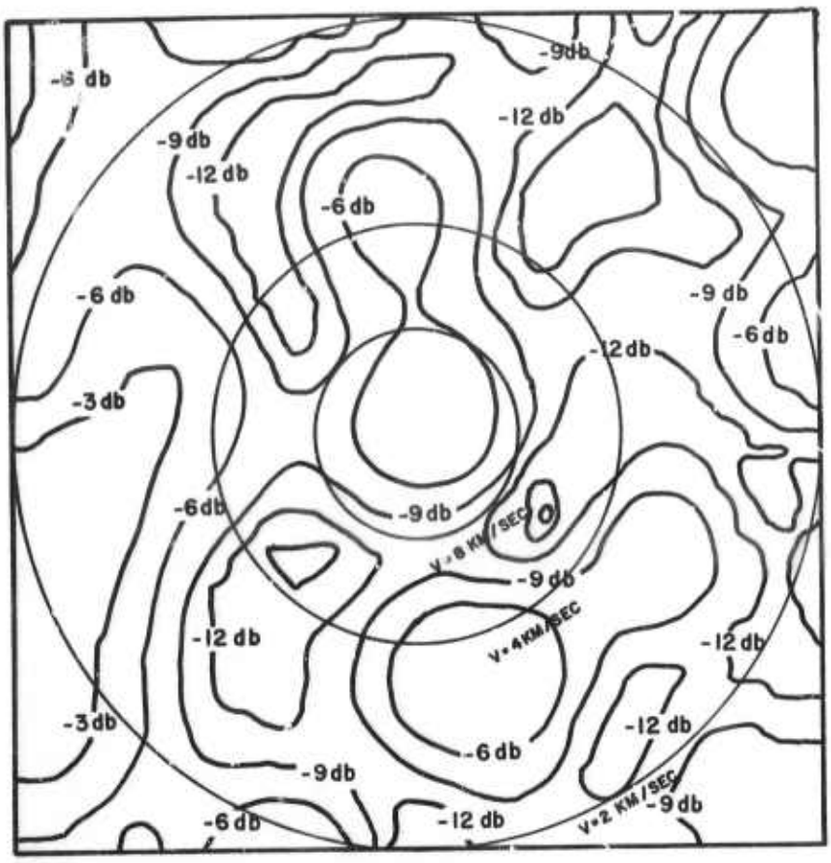


Figure 9b. FK Response of the Theoretical Isotropic Processor. Infinite Velocity Signal Model.

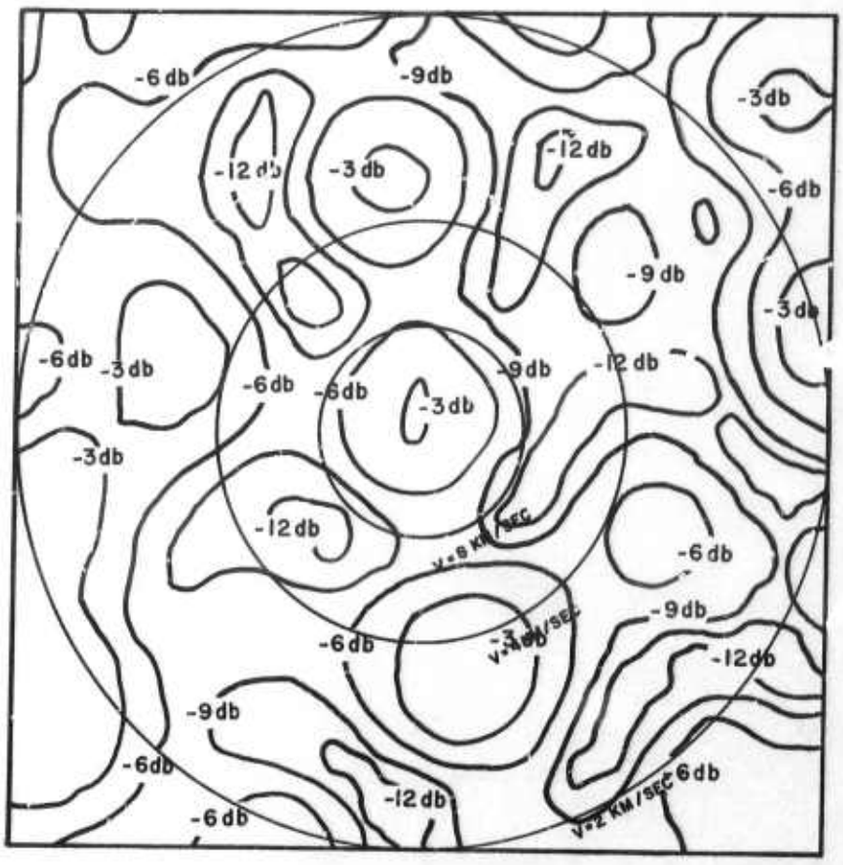
0.625 cps



1.25 cps

Figure 9c. FK Response of the Maximum Likelihood Filter. Designed on Un-Prefiltered Data.

0.625 cps



1.25 cps

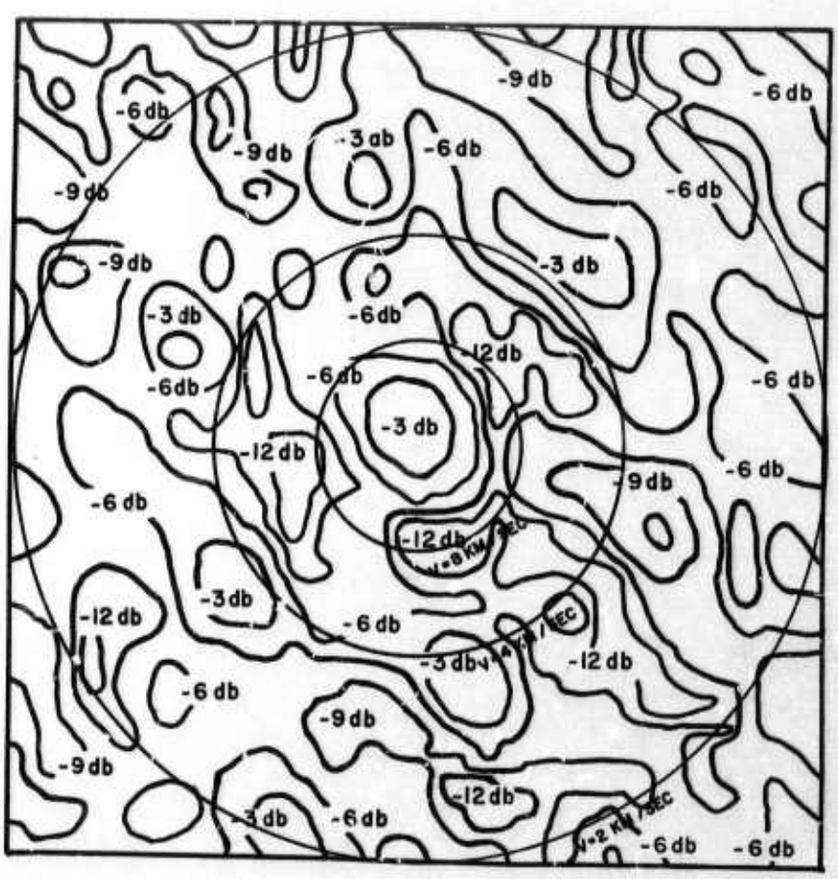
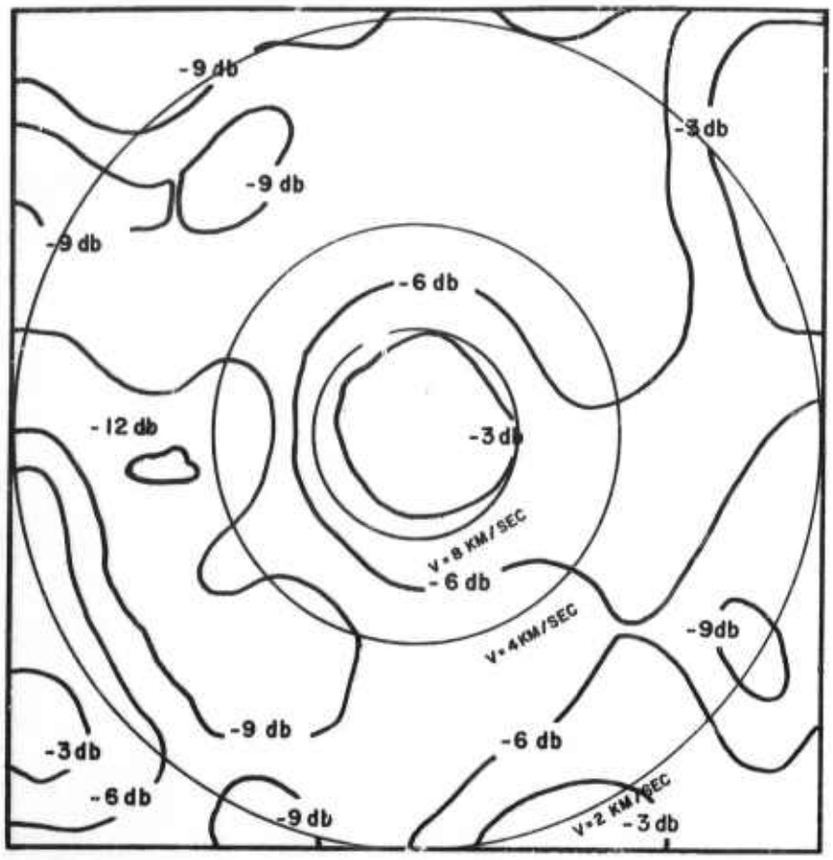


Figure 9d. FK Response of the Maximum Likelihood Filter. Designed on Prefiltered Data.

0.625 cps



1.25 cps

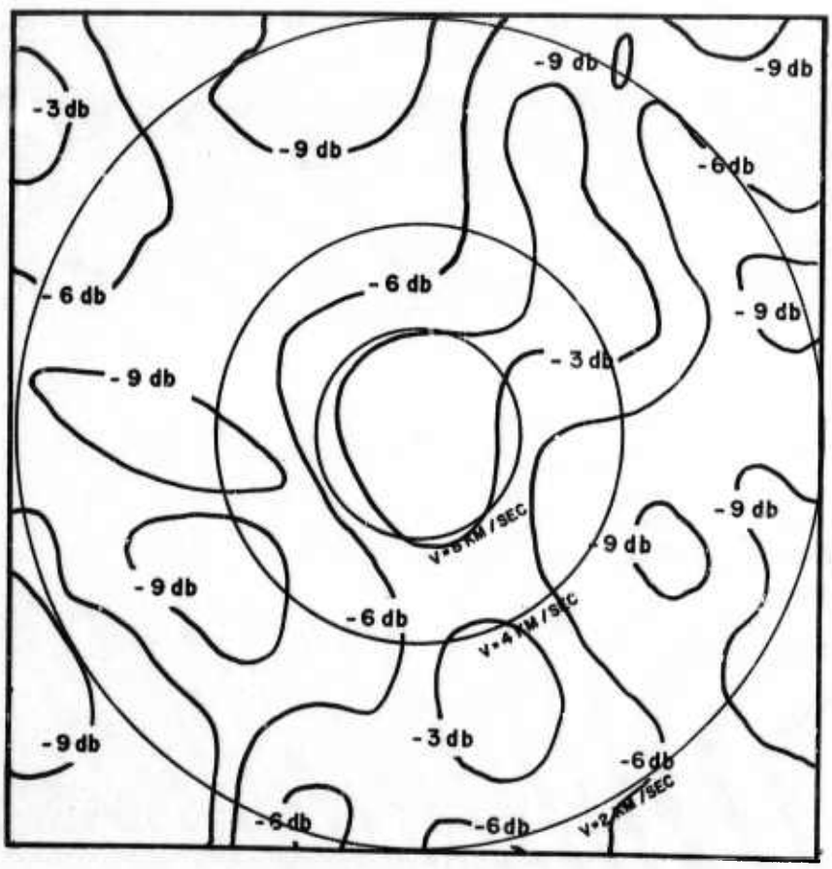
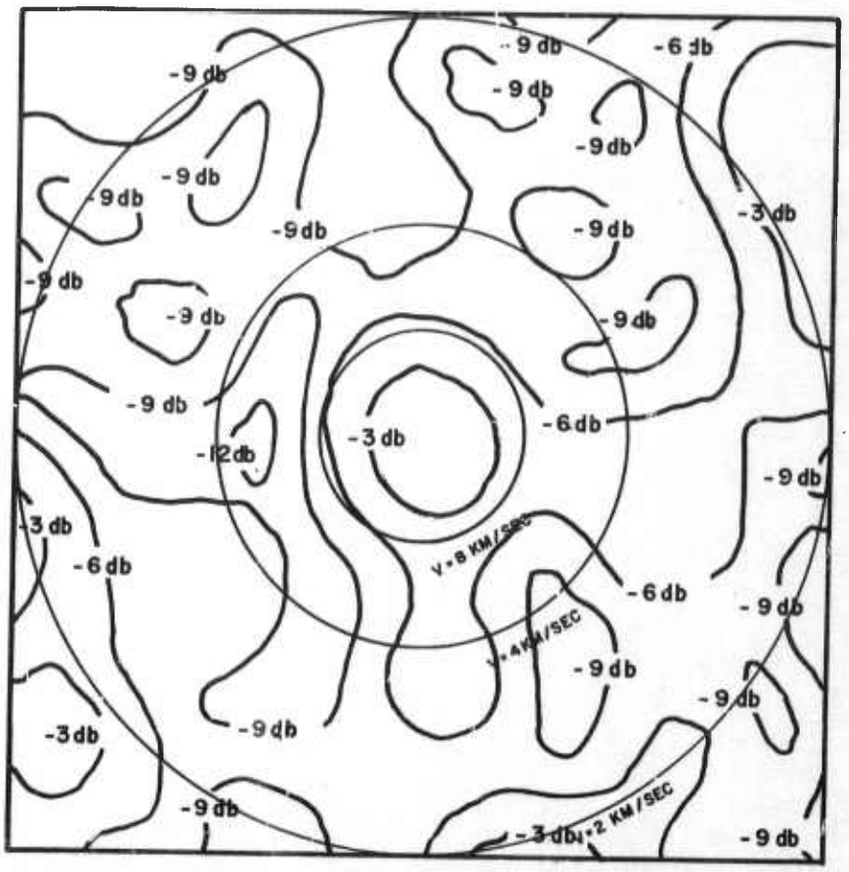


Figure 9e. FK Response of the Measured-Noise Isotropic Processor. Designed on Un-Prefiltered Data. 10 km/sec Signal Model.

0.625 cps



1.25 cps

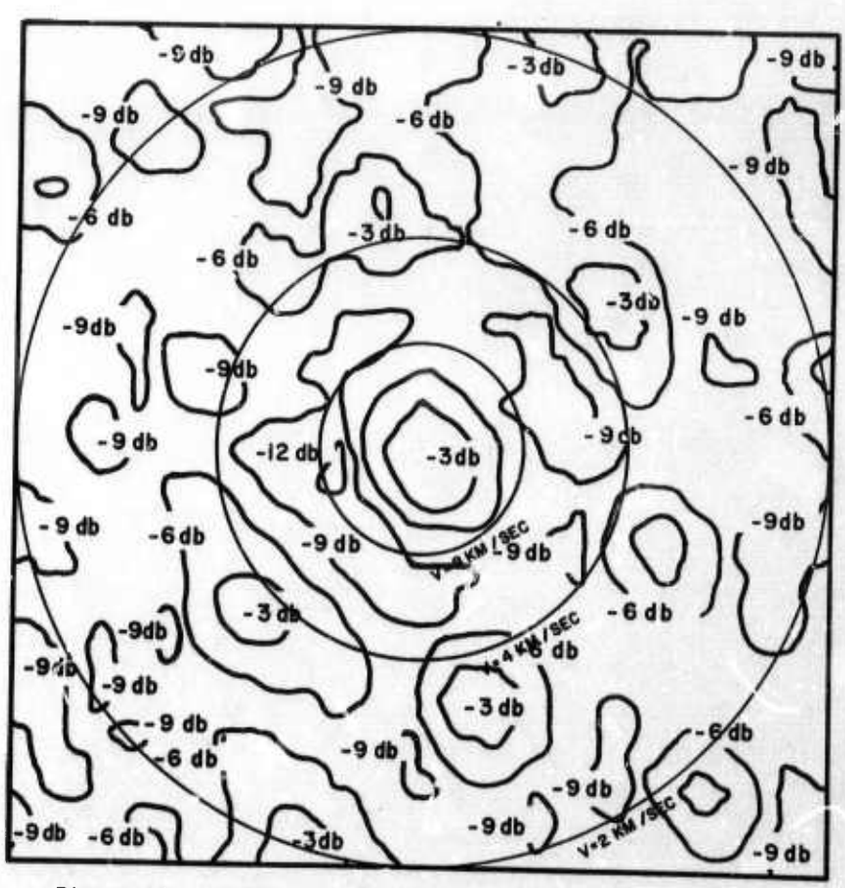
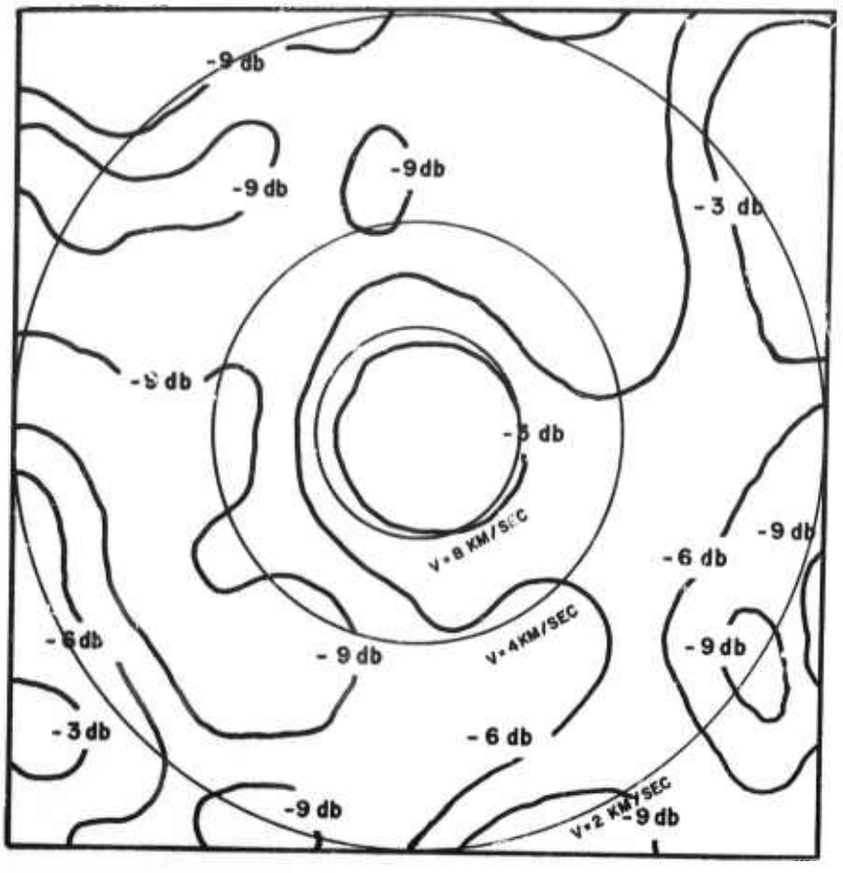


Figure 9f. FK Response of the Measured-Noise Isotropic Processor. Designed on Un-Prefiltered Data. Infinite Velocity Signal Model.

0.625 cps



1.25 cps

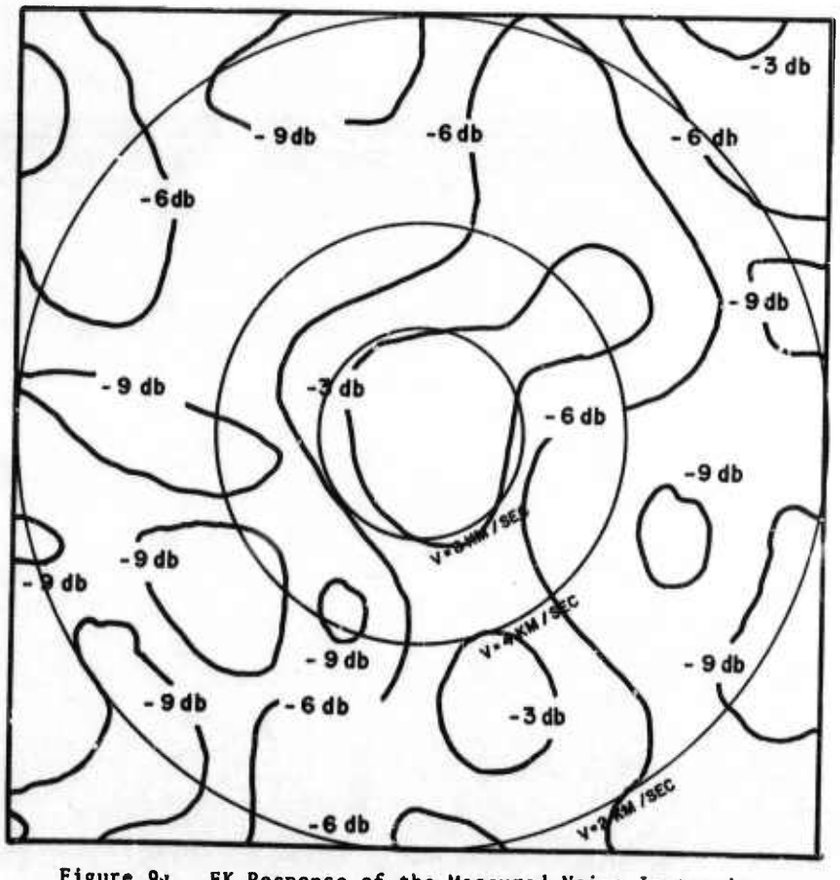
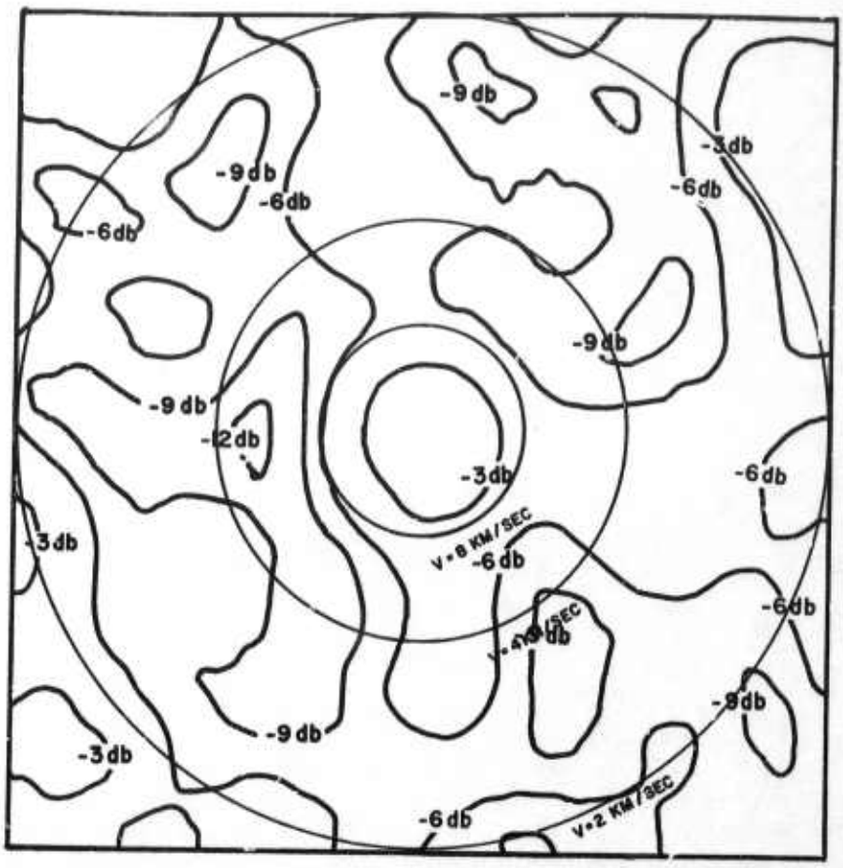


Figure 9g. FK Response of the Measured-Noise Isotropic Processor. Designed on Prefiltered Data. 10 km/sec Signal Model.

0.625 cps



1.25 cps

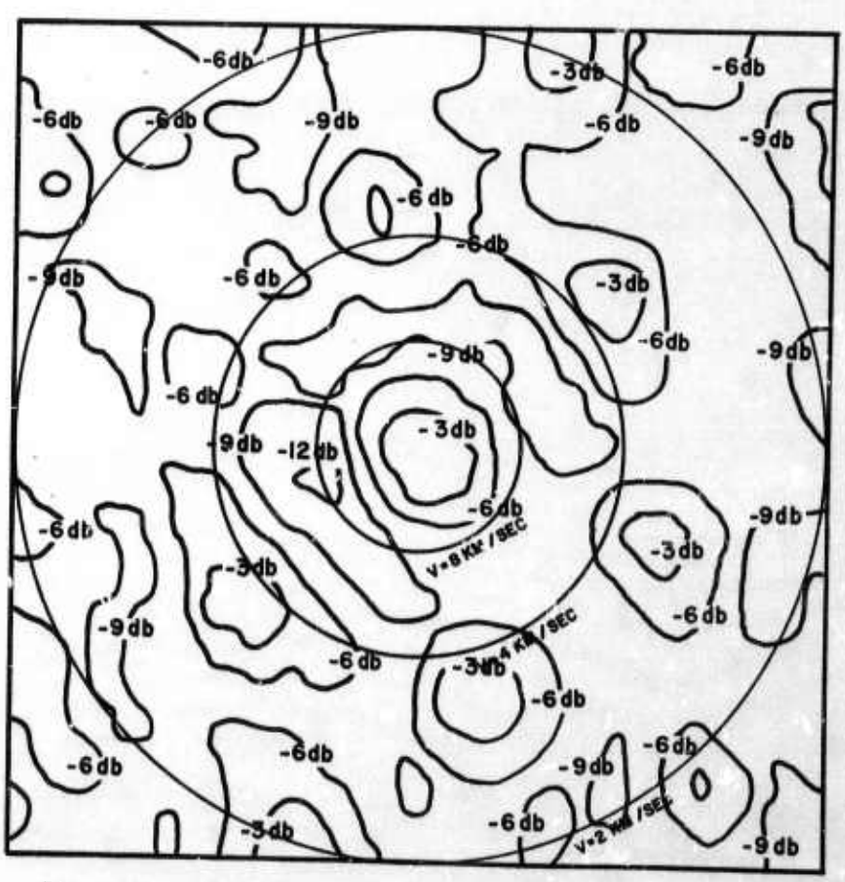
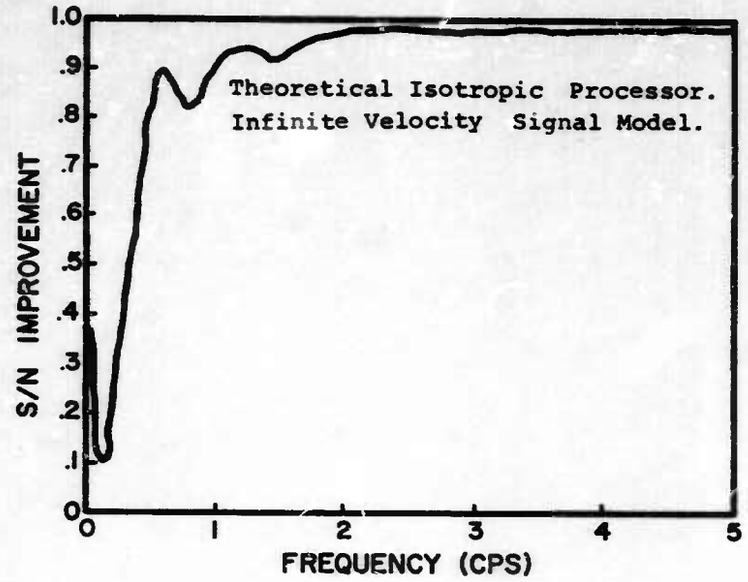
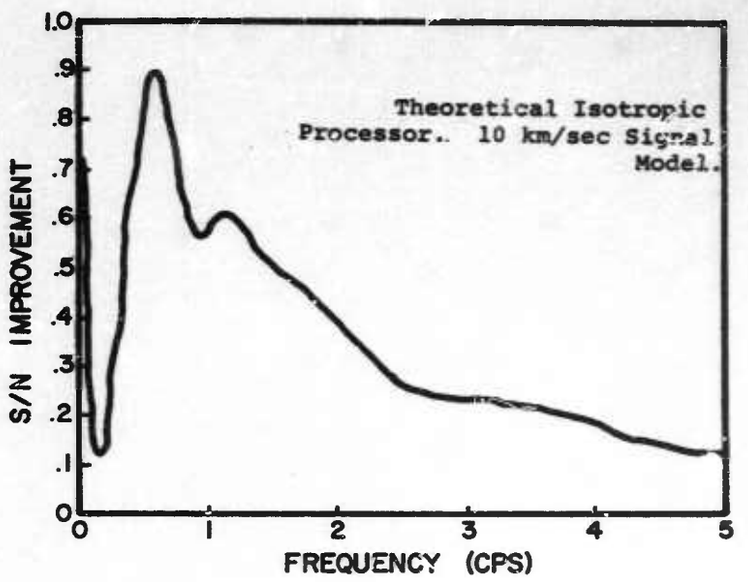
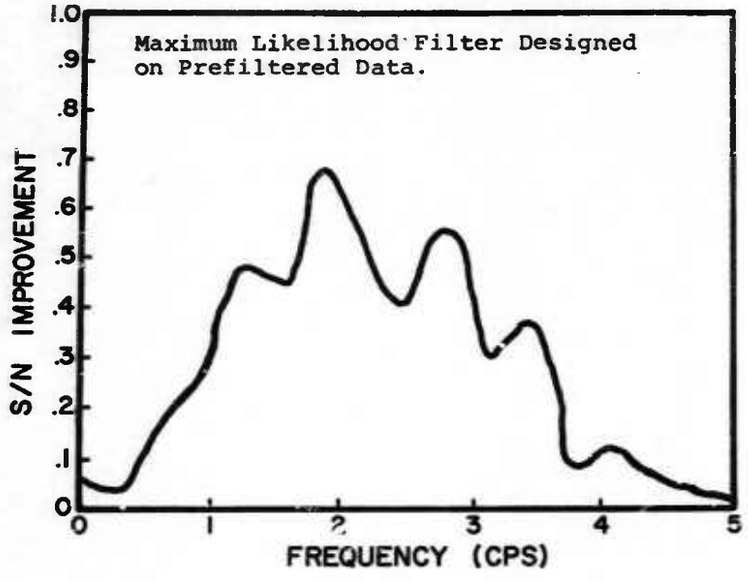


Figure 9h. MCF Response of the Measured-Noise Isotropic Processor. Designed on Prefiltered Data. Infinite Velocity Signal Model.







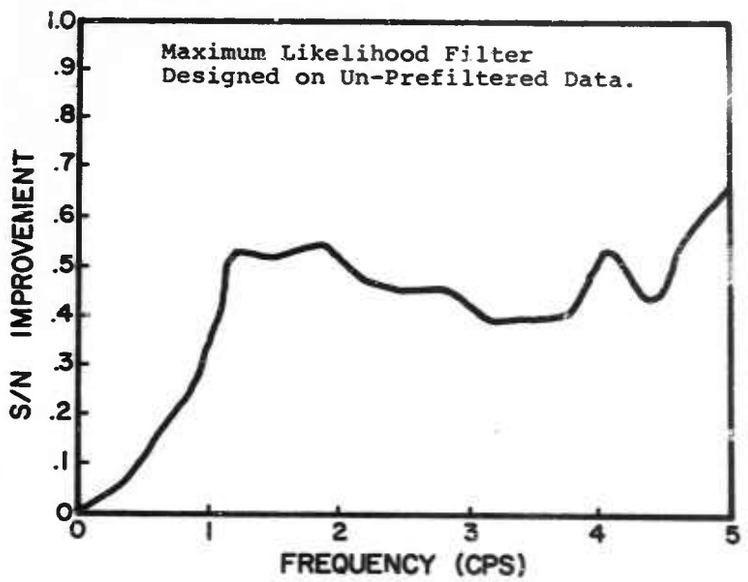
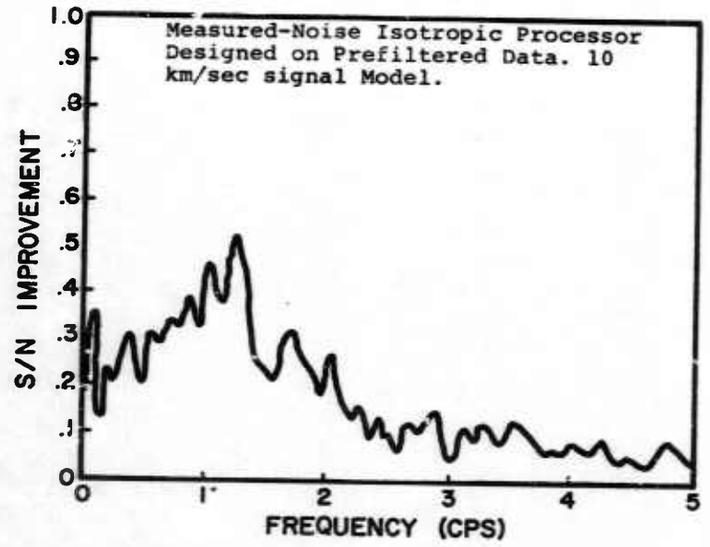
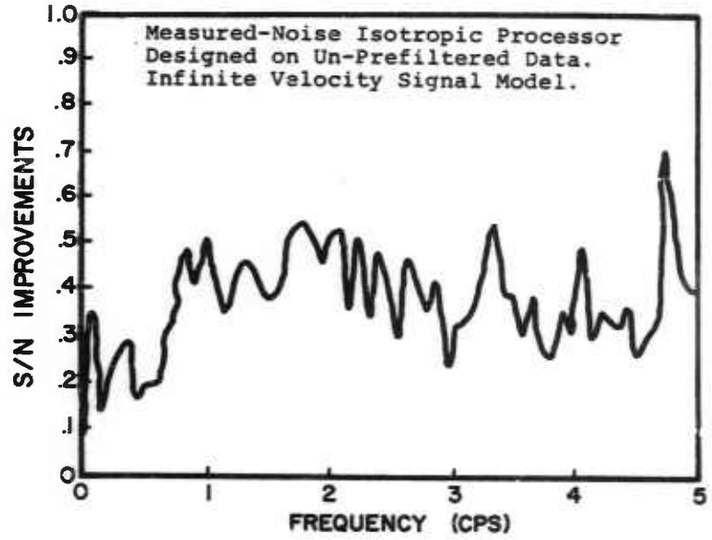
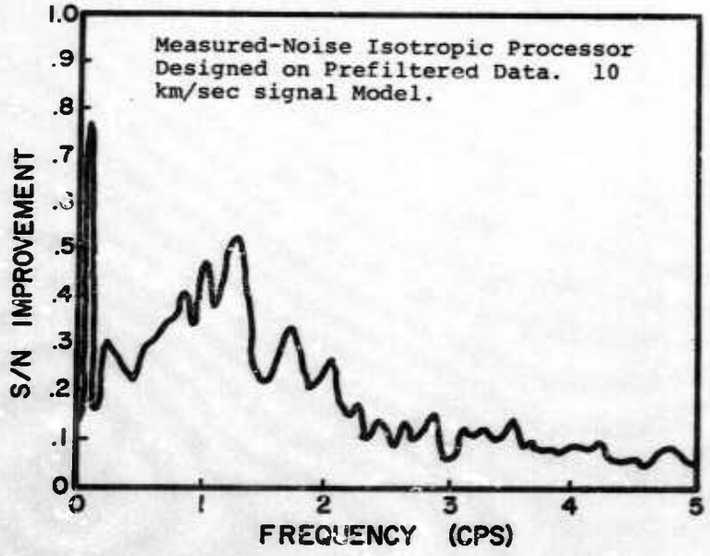


Figure 10a. Theoretical Signal to Noise Ratio Improvement vs Frequency for Multi-Channel Filters.







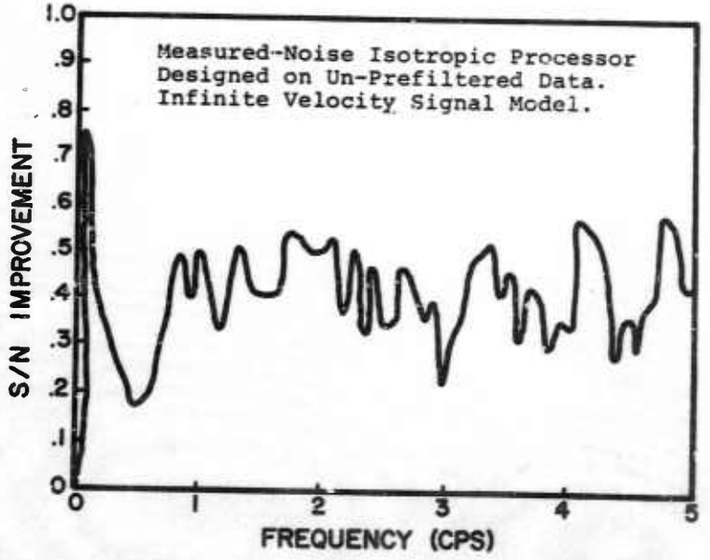


Figure 10b. Theoretical Signal to Noise Ratio Improvement vs Frequency for Multi-Channel Filters.

TABLE I

MULTICHANNEL FILTER REPORT

Location of Event

Date: 10 November 1965

Region: Aleutian Islands

Magnitude*: 4.1

Origin Time*: 03:48:78.0 GMT

Latitude*: 51.2° N

Longitude*: 178.8° W

Focal Depth*: "33 km"

Epicentral Distance to LASA A010: 46.1° (5126 km)

Azimuth of Epicenter from LASA A010: 304°

Surface Velocity at LASA A010: 14 km/sec

*Data from USC&S Preliminary Determination of Epicenter.

TABLE 2

MULTICHANNEL FILTER REPORT

Array Configuration

Channel No.	Channel ID	X Coordinate (km)	Y Coordinate (km)	Demagnification (counts/10^{-4} microns)
1	10	0.000	0.000	2.55
2	81	-0.716	3.426	3.09
3	32	0.745	0.667	2.78
4	52	1.610	1.440	3.18
5	83	3.325	-1.093	3.07
6	34	0.204	-0.979	2.48
7	54	0.409	-1.958	3.00
8	85	-2.609	-2.333	3.05
9	36	-0.950	0.312	2.69
10	56	-1.900	0.625	2.75

TABLE 3

MULTICHANNEL FILTER REPORT

S/N Ratio Improvement

Filter	Data	Signal Model	db Improvement
Unphased Sum	UBPF	--	0
Theoretical MCF	UBPF	10 km/sec	7
Theoretical MCF	UBPF	Infinite	4.8
Maximum Likelihood	UBPF	--	15
Measured-Noise MCF	UBPF	10 km/sec	3
Measured-Noise MCF	UBPF	Infinite	-6
Unphased Sum	BPF	--	15
Theoretical MCF	BPF	10 km/sec	16.4
Theoretical MCF	BPF	Infinite	16.1
Maximum Likelihood	BPF	--	17
Measured-Noise MCF	BPF	10 km/sec	10.3
Measured-Noise MCF	BPF	Infinite	5

Unclassified

Security Classification

DOCUMENT CONTROL DATA - R&D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author)
TELEDYNE, INC.
ALEXANDRIA, VIRGINIA

2a. REPORT SECURITY CLASSIFICATION
Unclassified

2b. GROUP

3. REPORT TITLE
DESIGN AND EVALUATION OF CERTAIN MULTICHANNEL FILTERS

4. DESCRIPTIVE NOTES (Type of report and inclusive dates)
Scientific

5. AUTHOR(S) (Last name, first name, initial)
McCowan, D. W.

6. REPORT DATE
26 January 1968

7a. TOTAL NO. OF PAGES
55

7b. NO. OF REFS
13

8a. CONTRACT OR GRANT NO.
F 33657-67-C-1313

b. PROJECT NO.
VELA T/6702

c.
ARPA Order 624

d. ARPA Program Code No. 5810

9a. ORIGINATOR'S REPORT NUMBER(S)
209

9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report)

10. AVAILABILITY/LIMITATION NOTICES
This document is subject to special export controls and each transmittal to foreign governments or foreign nationals may be made only with prior approval of Chief, AFTAC

11. SUPPLEMENTARY NOTES

12. SPONSORING MILITARY ACTIVITY
ADVANCED RESEARCH PROJECTS AGENCY
NUCLEAR TEST DETECTION OFFICE
WASHINGTON, D. C.

13. ABSTRACT

The theory of optimum least-squares multichannel filters is developed from the definitions of correlation functions and power spectra. Four types of these optimum processors were applied to an Aleutian earthquake recorded by part of the Montana LASA and the results evaluated. It appears that certain isotropic processors can give a modest signal-to-noise ratio improvement in excess of 1 db over the prefiltered straight sum. Evaluation of the outputs and transfer functions of these filters gives some insight into methods for improving their gain and stability. Several other topics are discussed from a theoretical point of view; these include estimation of power spectral density functions, frequency-wavenumber spectra, and certain coherence functions.

14. KEY WORDS

Digital Processing of Seismic Data
Multichannel Filters
Time Series Analysis
Coherence Analysis
Frequency-Wavenumber Spectra

Unclassified

Security Classification